Chemical Evolution II:
From the Origins of Life to Modern Society

Many a novice chemist will recall being asked to provide
a synthesis of target X starting from earth (C, Fe, S),
air (O_2, H_2, Cl_2), fire (Δ), and water (H_2O).

EDITED BY
Lori Zaikowski, Jon M. Friedrich
and S. Russell Seidel

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.fw001

Chemical Evolution II: From the Origins of Life to Modern Society

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.fw001

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.fw001

ACS SYMPOSIUM SERIES **1025**

Chemical Evolution II: From the Origins of Life to Modern Society

Lori Zaikowski, Editor
Dowling College

Jon M. Friedrich, Editor
Fordham University

S. Russell Seidel, Editor
Dowling College

Sponsored by the
ACS Division of Chemical Education

American Chemical Society, Washington DC

Library of Congress Cataloging-in-Publication Data

Chemical evolution II : from the origins of life to modern society / Lori Zaikowski, Jon M. Friedrich, S. Russell Seidel, editors ; sponsored by the ACS Division of Chemical Education.

p. cm. -- (ACS symposium series ; 1025)

Includes bibliographical references and index.

ISBN 978-0-8412-6980-4 (alk. paper)

1. Biochemistry--Congresses. 2. Life--Origin--Congresses. 3. Molecular evolution--Congresses. 4. Geochemistry--Congresses. 5. Chemistry, Technical--Congresses. I. Zaikowski, Lori, 1964- II. Friedrich, Jon M. III. Seidel, S. Russell. IV. Title: Chemical evolution 2.

QD415.A1C446 2009
572--dc22

2009031667

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.fw001

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48−1984.

Copyright © 2009 American Chemical Society

Distributed by Oxford University Press

All Rights Reserved. Reprographic copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Act is allowed for internal use only, provided that a per-chapter fee of $40.25 plus $0.75 per page is paid to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. Republication or reproduction for sale of pages in this book is permitted only under license from ACS. Direct these and other permission requests to ACS Copyright Office, Publications Division, 1155 16th Street, N.W., Washington, DC 20036.

The citation of trade names and/or names of manufacturers in this publication is not to be construed as an endorsement or as approval by ACS of the commercial products or services referenced herein; nor should the mere reference herein to any drawing, specification, chemical process, or other data be regarded as a license or as a conveyance of any right or permission to the holder, reader, or any other person or corporation, to manufacture, reproduce, use, or sell any patented invention or copyrighted work that may in any way be related thereto. Registered names, trademarks, etc., used in this publication, even without specific indication thereof, are not to be considered unprotected by law.

PRINTED IN THE UNITED STATES OF AMERICA

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.fw001

Foreword

The ACS Symposium Series was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of the series is to publish timely, comprehensive books developed from the ACS sponsored symposia based on current scientific research. Occasionally, books are developed from symposia sponsored by other organizations when the topic is of keen interest to the chemistry audience.

Before agreeing to publish a book, the proposed table of contents is reviewed for appropriate and comprehensive coverage and for interest to the audience. Some papers may be excluded to better focus the book; others may be added to provide comprehensiveness. When appropriate, overview or introductory chapters are added. Drafts of chapters are peer-reviewed prior to final acceptance or rejection, and manuscripts are prepared in camera-ready format.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previous published papers are not accepted.

ACS Books Department

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.pr001

Preface

Lori Zaikowski[1], Jon M. Friedrich[2], and S. Russell Seidel[1]

**[1]Department of Chemistry and Physics, Dowling College,
Oakdale, NY 11769
[2] Department of Chemistry, Fordham University,
Bronx, NY 10458**

In the first volume in this series, *Chemical Evolution across Space & Time* (*1*), we saw how the concept of evolution could be applied to chemical systems beginning with the origin of the most basic elements H, He, and Li in the Big Bang. Increasing complexity in the form of nucleosynthesis in stars and reactions around them followed pathways toward the synthesis of simple chemical compounds under the control of the basic principles of chemical binding and thermodynamics. The formation of our solar system led to more diverse chemical systems in the form of planetary atmospheres and the rocky parts of the planets and planetoids that they surround. This produced increasing chemical complexity coupled with an increasing diversity of chemical structures. A majority of *Chemical Evolution across Space & Time* dealt with descriptive investigations of what nature has created: humans acting as chemical archaeologists trying to comprehend the chemical history of the universe.

In this volume, a descriptive, investigative theme continues with humans using chemical principles to better understand the origins and evolution of life and the Earth's environment as we know it. Halfway through the volume, the perspective shifts. We begin to see how an increasing mastery of chemical construction – literally atom-by-atom – by humans can be used to create a deeper knowledge of biological and chemical processes. Hoye and Sizova point out that synthetic chemistry evolves, but not independently: advances in synthetic methods have changed chemistry. We are able to construct chemical structures that allow us to better examine nature's complex methodology. An additional underlying theme in the second half of the book is an increasing desire for humans to mimic, and attempt to perfect, nature's clever mastery of creating exceedingly tiny mechanically or electrically functional structures. Where the chemistry of the future lies is still a mystery and only time will tell. However, we know that the promise of human ingenuity and nanotechnology exemplified in this volume will continue to provide new applications and allow us increasing control of the produced chemical structures. The real question is what the effect of the structures synthesized with this control will be on the environment as a whole. As seen in the concluding chapter of this volume, the effects of human activity on evolutionary processes can be unexpected. Human ingenuity – and perhaps hubris – will continue to use the ideas of evolution as they relate to chemistry to better understand, appreciate and manipulate the chemical world.

This second volume on Chemical Evolution presents review chapters based on the proceedings of the symposium, "*Chemical Evolution II: From Origins of Life to Modern Society*", that was held on April 7-8, 2008 at the 235th National Meeting of the American Chemical Society in New Orleans, Louisiana. We sincerely thank the authors for their valuable contributions to the symposium and the book.

References

1. *Chemical Evolution across Space and Time: From the Big Bang to Prebiotic Chemistry;* Zaikowski, L.; Friedrich J. M., Eds.; American Chemical Society: Washington, DC, 2008; p 430.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.pr001

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

Chapter 1

Chemical Evolution: An Introduction

Robert M. Hazen

**Geophysical Laboratory, Carnegie Institution of Washington,
5251 Broad Branch Road NW, Washington DC 20015, USA**

Evolution, the process by which systems under non-random selective pressure become more complex, has been a pervasive force over 4.5 billion years of Earth history. Evolutionary episodes, including the prebiotic synthesis of biomolecules, the selection and organization of those molecules into self-replicating systems, the subsequent co-evolution of the geo- and biospheres, and the modern-day acceleration of chemical evolution through human invention, serve to illustrate this recurrent theme.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

Introduction

Complex evolving systems are a hallmark of the natural world and are observed in a variety of natural and human contexts, including nucleosynthesis in stars, diversification of minerals on terrestrial planets, prebiotic organic synthesis, development of languages, progress in material culture, and the biological evolution of life (*1-4*). In each of these complex systems the distribution of "species" evolves through non-random selective mechanisms, and each system displays such qualitatively similar characteristics as diversification into niches (radiation), episodic periods of innovation (punctuation), and the loss and replacement of previous species (extinction). However, these systems differ from each other in fundamental respects, notably in the degree to which their species demonstrate mutability, heritability and lateral transfer of traits. Comparisons among these disparate evolving systems thus point to general principles of emergent complexity.

Of all the chemical transformations that have shaped the Cosmos since its origin 14 billion years ago, none is more remarkable than the chemical origins of life. How did the raw materials of primitive Earth – oceans, atmosphere, rocks and minerals, and a diverse inventory of organic molecules – become alive? Details of that chemical history continue to foster intense debate, but the overarching narrative, as represented in the collected essays of these proceedings, is becoming clear.

This volume is the second of two that explore epic chemical changes in the Cosmos. The first volume (*1*) considered the emergence of chemical complexity from the earliest era of stellar evolution and the consequent emergence of elemental diversity, through planetary evolution and the emergence of mineralogical diversity, and ultimately to prebiotic chemical complexification and the emergence of life. Here the focus shifts to chemical and biological evolution on Earth. The story begins with the prebiotic origin of biomolecules, progresses to processes by which those molecules were selected and organized into self-replicating systems, examines the co-evolution of the geo- and biospheres, and culminates in human manipulation of the chemical realm.

The Building Blocks of Life

The first crucial step in life's emergence on Earth was the synthesis and gradual accumulation of abundant carbon-based biomolecules. In the beginning, life's raw materials consisted of oceans, an atmosphere of simple volcanic gases, varied rocks and minerals, and an expanding inventory of organic molecules that inexorably emerged from cosmochemical and geochemical environments. To understand life's origins we must decipher the chemical processes by which these basic ingredients reacted and complexified.

By the first decades of the 20th century experts agreed that life's chemical origins, wherever and however the processes occurred, depended on three key resources. First, water is the medium of all known life forms. All living cells,

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

even those that survive in the most extreme desert ecosystems, are formed largely of water. Consequently, the first cells are assumed to have arisen in an aqueous environment. The second requirement for life is a reliable energy source. The Sun's radiation provides the most obvious supply for life today, but bolts of lightning, impacts of asteroids, Earth's inner heat, and the chemical energy of minerals have also been invoked as life-triggering energy sources. Life's third requirement is an inventory of carbon, oxygen, hydrogen, and nitrogen, with lesser amounts of sulfur, phosphorus, and other elements as well.

The first step in life's chemical origins must have been the synthesis and accumulation of abundant biomolecules (*5-7*). Cleaves and Lazcano (*this volume*) review the rich variety of geochemical and astrochemical environments that promote synthesis of amino acids, carbohydrates, purines, pyrimidines, and a wide variety of other organic compounds. The experimental pursuit of geochemical organic synthesis, arguably the best understood aspect of life's origin, began in the early 1950s with the pioneering studies of University of Chicago graduate student Stanley Miller and his distinguished mentor Harold Urey (*8,9*). Together they demonstrated rapid organic synthesis triggered by the ionizing effects of simulated lightning. Other experiments using ultraviolet radiation or alpha particles revealed similar chemical synthetic pathways (*10,11*). So facile were these reactions that some researchers assumed that the problem of prebiotic chemical evolution had been solved.

The discovery of abundant organic molecules in some carbonaceous meteorites points to additional important extraterrestrial sources of organic molecules (*12-14*). Reactions of UV-irradiated small molecules in the cold vacuum of interstellar space produce complex suites of organics (*15,16*) – processes that have been replicated in laboratory experiments (*17,18*). Comets and asteroids continuously delivered supplies of these molecules to the early Earth.

Lightning, alpha particles, and ultraviolet radiation promote chemical reactions through ionization and the formation of energetic free radicals. Living cells, on the other hand, build organic molecules through sequences of less-energetic metabolic redox reactions. Many origins-of-life researchers therefore point to natural geochemical environments, including deep-ocean hydrothermal vents and reactive surfaces of transition-metal minerals, as equally plausible energy sources for prebiotic synthesis (*19-22*). Astrobiologists have strong incentives for examining the possibility of such mineral-mediated deep origins. If life is constrained to form in a warm little pond or ocean surface then Earth, and perhaps ancient Mars or Venus, are the only possible places where life could have begun in our Solar System. However, if cells can emerge from deeply buried wet zones, then Europa, Callisto, Titan, and other bodies become promising targets for astrobiological exploration.

Russell and Hall (*this volume*) adopt this premise and suggest that life first emerged where dissolved H_2 from submarine hot springs reduced dissolved CO_2 in the ancient ocean. They propose that iron-nickel sulfide minerals served as catalysts in the generation of essential biomolecules. Other important thermodynamic energy sources, Russell and Hall note, resulted from the strong pH and thermal gradients between the acidic ocean and alkaline hydrothermal fluids.

Amend and McCollom (*this volume*) echo and amplify this view by calculating thermochemical parameters of redox reactions that must have constituted the earliest metabolisms at geochemically relevant temperatures, pressures, and chemical compositions. They thus demonstrate that the molecular components that make up a prokaryotic cell can be produced through aqueous synthesis from inorganic precursors.

The conclusion of more than half a century of research is that prebiotic synthesis of biomolecules took place in numerous astrochemical and geochemical environments. Consequently, Earth's primitive oceans must have been a complex, if dilute, broth of organic chemicals. Regardless of their synthetic sources, these simple organic compounds, so widely dispersed in the primitive oceans, had to be concentrated and organized by simple physicochemical mechanisms.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

Molecular Selection and Organization

The abundant synthesis of varied prebiotic organic molecules was a prerequisite for life, and it is satisfying that so many plausible mechanisms for their formation have been demonstrated experimentally. However, this profligate production raises another question: how were the essential molecules of life selected, concentrated and organized from the messy prebiotic soup. The authors of this volume demonstrate a range of possibilities.

The oceans are of little help because they are so vast – a volume greater than a billion cubic kilometers. Even given the most optimistic estimates of prebiotic organic synthesis (*5*), the oceans were never more than a hopelessly dilute solution. Making biological macromolecules is further complicated by the fact that for every potentially useful small molecule in the prebiotic milieu, dozens of other molecular species had no obvious role in biology. Life is remarkably selective in its building blocks; the vast majority of carbon-based molecules synthesized in prebiotic processes have no biological use whatsoever. In such a random, weak solution it would have been difficult for just the right combination of molecules to bump into one another to make anything useful in the chemical path to life.

Consider sugar molecules, for example (*23*). All living cells rely on two species of 5-carbon sugar molecules, ribose and deoxyribose (the "R" and "D" in RNA and DNA, respectively). Several plausible prebiotic synthesis pathways yield a small amount of these essential sugars, but for every ribose molecule produced many other kinds of 5-carbon sugar also appear – xylose, arabinose, and lyxose, for example. Adding to this chemical complexity are the numerous 3-, 4-, 6- and 7-carbon sugars, in chain, branch and ring structures. Furthermore, many sugar molecules, including ribose and deoxyribose, come in mirror-related pairs – left- and right-handed varieties that possess the same chemical formula and many of the same physical properties, but differ in shape like left and right hands. Prebiotic synthesis pathways generally yield equal amounts of left- and right-handed sugars, but cells employ only the right-handed sugar varieties. Consequently, many origins researchers focus on the processes by which molecules might have been selected, concentrated and organized into the essential structures of life.

Molecular self-organization represents one solution to the problem, as illustrated by the behavior of phospholipid molecules. These long, slender building blocks of cell membranes feature one hydrophilic (water loving) end, while the rest of the molecule is hydrophobic (water hating). Consequently, when placed in water, life's lipids spontaneously self-organize into tiny cell-like spheres – a self-organizing process that is rapid and spontaneous (*24-26*).

Zimmerman *et al.* (*this volume*) demonstrate an example of this self-selection process in the base-pairing of adenine and thymine (the A and T of the genetic alphabet), which together display unexpected stability. Zimmerman and colleagues employed both experiments and computations to compare the free energy of formation of numerous base pairs. The exceptional results for the A-T combination suggest that this pairing may represent a deterministic aspect of life's chemical origins.

Continuing on this theme, Seidel and Zaikowski (*this volume*) review examples of coordination-driven self-assembly of macromolecular structures, including a variety of polyhedral forms, as well as large rings and cages. The observed processes of supramolecular assembly have direct parallels to self-assembly of structures in biological systems. Similar self-assembly phenomena must also have played a significant role in the origin of the first RNA-like polymers, according to Bean, Lynn, and Hud (*this volume*). They hypothesize that life's first informational polymers incorporated chemical modules similar to those of RNA, but were distinct in that they self-organized in prebiotic environments. Furthermore, they suggest, small catalytic peptides formed through drying-heating reactions and assembled into surfaces and vesicles. This postulated co-evolution of peptide enzymes and genetic polymers represents one viable compromise solution in the chicken-and-egg dilemma of whether a metabolic cycle or genetic polymers came first in life's origins.

In spite of these advances, most molecules don't self organize. Consequently, many scientists have focused on surfaces as an alternative solution to the problem of selection and organization. Chemical complexity often arises at surfaces, where different molecules can congregate and interact. The surface of the ocean where air meets water is one promising interface, where a primordial oil slick might have concentrated organic molecules (*27*). Evaporating tidal pools where rock and water meet and cycles of evaporation concentrate stranded chemicals provide another appealing scenario for origin-of-life chemistry (*28*). Deep within the crust and in hydrothermal volcanic zones mineral surfaces may have embraced a similar role, selecting, concentrating and organizing molecules on their periodic crystalline surfaces (*29-32*).

Solid rocks provide especially attractive surfaces for concentration and assembly of molecules. Experiments reveal that amino acids concentrate and polymerize on clay particles to form small, protein-like molecules, while layered minerals also have the ability to adsorb and assemble the building blocks of RNA or accumulate small organic molecules in the relatively large spaces between layers (*33-35*). Once confined and concentrated, these small molecules tend to react to form larger molecular species that aren't otherwise likely to emerge from the soup.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

One of the most intriguing and confounding examples of prebiotic molecular selection is the ancient incorporation of handedness. Many of the most important biomolecules, amino acids and sugars included, come in mirror image “chiral” pairs. These left- and right-handed molecules have virtually the same energies and physical properties, and all known prebiotic synthesis pathways produce chiral molecules in essentially 50:50 mixtures. Thus, no obvious inherent reason exists why left or right should be preferred, yet living cells display the most exquisite selectivity, choosing right-handed sugars over left, and left-handed amino acids over right (*7,32,36,37*).

Some analyses of chiral amino acids in carbonaceous meteorites point to the possibility that Earth was seeded by amino acids that already possessed a left-handed bias (*38-40*), though recent analyses of amino-acid-rich meteorites may cast doubt on these findings (*14*). According to one scenario, left-handed molecules could have been concentrated if circularly polarized synchrotron light from a rapidly rotating neutron star selectively photolyzed right-handed amino acids in the solar nebula (*41-43*). However, it is also difficult to eliminate entirely the possibility of a left-handed overprint imposed in the laboratory during the difficult extraction and analysis of trace quantities of meteorite organics.

Alternatively, many origin-of-life researchers argue that the chirality of life occurred as a chance event – the result of an asymmetric local physical environment on Earth. Such local chiral environments abounded on the prebiotic Earth, both as chiral molecules, themselves, and in the form of asymmetric mineral surfaces (*44,45*). Minerals often display chiral crystal faces, which might have provided templates for the assembly of life’s molecules. Experiments show that left- and right-handed mineral surfaces provide one possible solution for separating a 50:50 mixture of L and D molecules (*46,47*).

Whatever the origin, either global or local, of chiral molecular excesses, the subsequent amplification of chirality has been successfully tackled by Klussmann and Blackmond (*this volume*). They provide an experimental demonstration of physical mechanisms based on the phase behavior of chiral amino acids in aqueous systems and in equilibrium, making it a robust and likely model for the prebiotic soup.

The emergence of highly selected and organized macromolecular structures is an essential step, but not the final one, in the progression from geochemistry to biochemistry. Life requires that macromolecules be incorporated into a self-replicating system. Two contrasting models of the first self-replicating system are debated. On the one hand, metabolism, by which chemicals react to release energy and manufacture molecules that reinforce the metabolic cycle, requires a sequence of chemical reactions that work in concert (*3,7,29,48*). On the other hand, the genetics-first scenario assumes a self-replicating and catalytic molecule like RNA that promotes its own replication while passing information from one generation to the next (*49-51*). Whichever scenario proves correct, a self-replicating molecular system, with its intrinsic potential for small chemical mutations and thus competition from competing cycles, would have experienced rapid evolution by the process of natural selection (*52,53*). This chemical origin of life forever transformed Earth’s surface and led to the co-evolution of the geo- and biospheres.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

Co-Evolution of the Geo- and Biospheres

Four billion years ago the seeds of life had been firmly planted. The Archean Earth boasted substantial repositories of serviceable organic molecules, which became locally concentrated and assembled into vesicles and polymers of biological interest. Once the first molecular replicator emerged, molecular natural selection took off. In such a world, nascent biology and ancient geochemistry became inextricably entwined. For example, Earth's surface mineralogy diversified as a result of varied microbial influences, which altered ocean and atmospheric chemistries at scales from local to global (*2*).

Authors in this volume then explore how the chemical environment at Earth's surface has evolved since life began. The greatest single change in Earth's 4.5 billion year history was the transition from an anoxic world to the oxygenated world of today – the byproduct of oxygenic photosynthesis, which greatly increased the energy available to the biosphere. We now recognize two major increments in atmospheric oxygen – the Great Oxygenation Event of 2.2 to 2.0 Ga (*54-56*) and the post-glacial Neoproterozoic events of 0.75 to 0.54 Ga (*57,58*). These transitions created profound biochemical challenges, because life had to develop new strategies to harvest redox-sensitive, bioessential elements such as Fe, Cu and Mo, while combating oxidative cellular damage. Burrows (*this volume*) explores how eukaryotes, which emerged about 2 billion years ago, initiated DNA repair to survive in the presence of reactive oxygen species. She reviews the remarkable evolution of DNA polymerase and repair enzymes that deal with the estimated thousands of daily oxidative damage events.

Continuing in this vein, Hemming (*this volume*) considers more recent feedbacks between life and atmospheric changes that affect climate. He uses boron isotopes, which serve as a proxy for ancient ocean pH. Because of interactions between the atmosphere and surface ocean, these isotopic data reveal natural variations in atmospheric CO_2 concentrations. Understanding such natural co-variability in atmospheric composition and climate is essential if we are to document and predict the role of anthropogenic influences on present and future climates.

In this volume's concluding contribution, Palumbi considers a troubling aspect of the intimate connection between the geo- and biospheres – that is, the rapid biological evolution that may result from human-induced environmental changes. Significant changes in the biosphere, including landscape fragmentation, overuse of antibiotics, application of pesticides, overfishing and climate change, exert strong bio-selection pressures. Palumbi explains that for species with large populations and rapid generation times, including many pathogens and insects, rapid evolution is often the result. However, species with relatively small populations and long reproductive cycles (polar bears and the great apes come to mind) often exhibit rapid population declines under such environmental changes. These evolutionary consequences of human-induced environmental changes are not abstract findings. The emergence of new highly-virulent anti-biotic-resistant pathogens and aggressive herbicide-resistant weeds has resulted in increased costs for medical care and agricultural products. However, such unintended consequences of human activities are only part of the modern story of chemical evolution.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

Directed Chemical Evolution

The past two centuries have seen the arrival of a new era of chemical evolution, one fostered by human invention and creativity. Hoye (*this volume*) reviews this rapid evolution of chemistry through developments in synthesis, especially organic synthesis. He emphasizes how progress has been accelerated through the use of theoretical models, analytical developments, and a host of novel catalysts.

A number of challenging chemical problems serve to illustrate advances in the modern era of directed chemical evolution. Muckerman and Fujita (*this volume*) consider efforts to develop artificial photosynthesis, by which sunlight, water and carbon dioxide are converted into carbohydrates and oxygen. They summarize two distinct approaches to realizing the goal of artificial photosynthesis: structural models vs. functional models of the natural systems. Shen and Rajski (*this volume*) describe a similar natural systems strategy to developing drugs that mimic biological products, which are the basis for most new drug discovery and development. He observes that advances in genetics and in documenting biosynthetic pathways are making it possible to identify valuable natural products and to craft new synthetic chemicals.

Marti and Turro (*this volume*) discuss the flourishing field of fluorescent responsive molecular probes for use in the detection of oligonucleotides. Varied strategies, including molecular beacons and binary probes, are now used to track mRNA *in vivo*, quantify the polymerase chain reaction *in vitro*, and many other applications. The implementation of dye combinations, excimer-forming molecules, and metal complexes is further expanding the range of applications for these fluorescent markers.

Given the vast number of possible useful chemicals, combinatorial approaches to chemical synthesis and testing are proving to be a valuable way to identify highly functional molecules. A step beyond this synthesis of individual chemical species is the nano-world of molecular machines, such as artificial molecular motors, which require the collective interactions of numerous specialized molecules. Bell (*this volume*) considers major types of natural and artificial molecular motors, with an emphasis on recent research to design energy-efficient molecular subunits.

Conclusions: Complex Evolving Systems

This volume outlines how three distinct stages of chemical evolution – prebiotic, biological, and human-directed – have framed Earth's 4.5-billion year history. Each stage saw significant chemical complexification driven by varied selection processes. In the prebiotic era equilibrium physico-chemical mechanisms prevailed, as atmosphere, oceans, and rocks interacted under a wide range of environments on and beneath Earth's solid surface. Synthesis was promoted by solar and geothermal energy sources, amplified by such geochemical complexities as thermal and compositional gradients, fluid fluxes, solid-fluid and fluid-fluid interfaces, and cycles of light, heat, and tides. These environmental variables promoted the synthesis, selection, concentration and

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

organization of organic molecules into the first self-replicating, mutable chemical entities – the first form of life.

Life, with its paired abilities to mutate and to pass those mutations on to subsequent generations, promoted our planet's second stage of chemical evolution and inexorably increased the variety and modified the distribution of chemical substances in Earth's near-surface environment. The innovation of catalytic proteins and polynucleotides, in particular, altered the chemical landscape of our planet. Consequently, for at least three billion years microbial activities have altered the chemistries of Earth's atmosphere and oceans, thus driving the co-evolution of the geo- and biospheres. Atmospheric oxygenation, the innovation of biomineralized skeletons, and the colonization of terrestrial habitats all expanded the influence of life on Earth's surface chemistry.

The past two centuries have seen a third, far more rapid stage of human-directed chemical evolution, with global-scale exploitation of minerals and fossil fuels, and the synthesis of countless new chemical products. These dramatic developments have resulted in remarkable improvements in human health and productivity, but changes in our chemical environment have also led to unintended consequences, both known and as yet unresolved. Let us hope that with our new ability to guide chemical evolution will also come the insight and desire to use that power wisely.

References

1. *Chemical Evolution Across Space and Time;* Zaikowski, L.; Friedrich, J., Eds.; ACS Symposium Series 981; American Chemical Society: Washington, DC, 2008.
2. Hazen, R. M.; Papineau, D.; Bleeker, W.; Downs, R. T.; Ferry, J.; McCoy, T. L.; Sverjensky, D. A.; Yang, H. *Am. Mineral.* **2008,** *93,* 1693-1720.
3. Morowitz, H. J. *The Emergence of Everything;* Oxford University Press: New York, 2002.
4. Chaisson, E. J. *Cosmic Evolution;* Harvard University Press: Cambridge, MA, 2001.
5. Chyba, C. F.; Sagan, C. *Nature* **1992,** *355,* 125–132.
6. Lahav, N. *Biogenesis;* Oxford University Press: New York, 1999.
7. Hazen, R. M. *Genesis: The Scientific Quest for Life's Origins;* Joseph Henry Press: Washington, DC, 2005.
8. Miller, S. L. *Science* **1953,** *17,* 528–529.
9. Miller, S. L.; Urey, H. C. *Science* **1959,** *130,* 245–251.
10. Garrison, W. M.; Morrison, D. C.; Hamilton, J. G.; Benson, A. A.; Calvin, M. *Science* **1951,** *114,* 416–418.
11. Abelson, P. H. *Proc. Natl. Acad. Sci. U.S.A.* **1966,** *55,* 1365–1372.
12. Kvenvolden, K. A.; Lawless, J. G.; Pering, K.; Peterson, E.; Flores, J.; Ponnamperuna, C.; Kaplan, I. R.; Moore, C. *Nature* **1970,** *28,* 923–928.
13. Deamer, D. W.; Pashley, R. M. *Origins Life Evol. Biosphere* **1989,** *19,* 21–38.
14. Martins, Z.; Alexander, C. M. O'D.; Orzechowska, G. E.; Fogel, M. E.; Ehrenfreund, P. *Meteorit. Planet. Sci.* **2008,** *42,* 2125–2136.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

15. Pendleton, Y. J.; Chiar, J. E. In *From Stardust to Planetesimals;* Pendleton, Y. J.; Tielens, A. G. G. M., Eds.; ASP Conference Proceedings Vol. 122; Astronomical Society of the Pacific: Provo, UT, 1997; pp 179–200.
16. Rawls, R. L. *Chem. Eng. News* July 15, 2002, pp 31–37.
17. Bernstein, M. P.; Dworkin, J. P.; Sandford, S. A.; Cooper, G. W.; Allamandola, L. J. *Nature* **2002,** *416,* 401–403.
18. Muñoz Caro, G. M.; Meierhenrich, U. J.; Schutte, W. A.; Barbier, B.; Segovia, A. A.; Rosenbauer, H.; Thiemann, W. H.-P.; Brack, A.; Greenberg, J. M. *Nature* **2002,** *416,* 403–406.
19. Blöchl, E.; Keller, M.; Wächtershäuser, G.; Stetter, K. O. *Proc. Natl. Acad. Sci. U.S.A.* **1992,** *89,* 8117–8120.
20. Heinen, W.; Lauwers, A. M. *Origins Life Evol. Biosphere* **1996,** *26,* 131–150.
21. Cody, G. D.; Boctor, N. Z.; Filley, T. R.; Hazen, R. M.; Scott, J. H.; Sharma, A.; Yoder, H. S., Jr. *Science* **2000,** *289,* 1337–1340.
22. Cody, G. D.; Boctor, N. Z.; Brandes, J. A.; Filley, T. R.; Hazen, R. M.; Yoder, H. S., Jr. *Geochim. Cosmochim. Acta* **2004,** *68,* 2185–2196.
23. Nelson, D. L.; Cox, M. M. *Lehninger's Principles of Biochemistry, 4th Edition;* Worth Publishers: New York, 2004.
24. Luisi, P. L.; Varela, F. J. *Origins Life Evol. Biosphere* **1989,** *19,* 633–643.
25. Deamer, D. W. *Microbiol. Mol. Biol. Rev.* **1997,** *61,* 239–261.
26. Segré, S.; Deamer, D. W.; Lancet, D. *Origins Life Evol. Biosphere* **2001,** *31,* 119–145.
27. Lasaga, A. C.; Holland, H. D.; Dwyer, M. J. *Science* **1971,** *174,* 53–55.
28. Lahav, N.; White, D.; Chang, S. *Science* **1978,** *201,* 67–69.
29. Wächtershäuser, G. *Prog. Biophys. Mol. Biol.* **1992,** *58,* 85–201.
30. Orgel, L. E. *Origins Life Evol. Biosphere* **1998,** *28,* 227–234.
31. Ferris, J. P. *Biology Bulletin* **1999,** *196,* 311–314.
32. Hazen, R. M.; Sholl, D. S. *Nat. Mater.* **2003,** *2,* 367–374.
33. Ferris, J. P. *Origins Life Evol. Biosphere* **1993,** *23,* 307–315.
34. Pitsch, S.; Eschenmoser, A.; Gedulin, B.; Hui, S.; Arrhenius, G. *Origins Life Evol. Biosphere* **1995,** *25,* 297–334.
35. Smith, J. V. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95,* 3370–3375.
36. Bonner, W. A. *Origins Life Evol. Biosphere* **1991,** *21,* 59–111.
37. Bonner, W. A. *Origins Life Evol. Biosphere* **1995,** *25,* 175–190.
38. Cronin, J. R.; Pizzarello, S. *Adv. Space Res.* **1983,** *3,* 5–18.
39. Engel, M. H.; Macko, S. A. *Nature* **1997,** *296,* 837–840.
40. Pizzarello, S.; Cronin, J. R. *Geochim. Cosmochim. Acta* **2000,** *64,* 329–338.
41. Clark, S. *American Scientist* **1999,** *87,* 336–343.
42. Bailey, J.; Chrysostomou, A.; Hough, J. H.; Gledhill, T. M.; McCall, A.; Clark, S.; Menard, F.; Tamura, M. *Science* **1998,** *281,* 672–674.
43. Podlech, J. *Angew. Chem. Int. Ed. Engl.* **1999,** *38,* 477–478.
44. Hazen, R. M. In *Progress in Biological Chirality;* Palyi, G.; Zucchi, C.; Caglioti, L., Eds.; Elsevier: Oxford, United Kingdom, 2004; pp 137–151.
45. Hazen, R. M. *Am. Mineral.* **2006,** *91,* 1715–1729.
46. Hazen, R. M.; Filley, T. R.; Goodfriend, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98,* 5487–5490.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

47. Castro-Puyana, M.; Salgado, A.; Hazen, R. M.; Crego, A. L.; Marina, M. L. *Electrophoresis* **2008,** *29,* 1548–1555.
48. Russell, M. J.; Hall, A. J. *J. Geol. Soc. London* **1997,** *154,* 377–402.
49. Orgel, L. E. *J. Theor. Biol.* **1986,** *123,* 127–149.
50. Joyce, G. F. *New Biology* **1991,** *3,* 399–407.
51. Orgel, L. E. *Origins Life Evol. Biosphere* **2003,** *33,* 211–218.
52. Eigen, M.; Schuster, P. *The Hypercycle: A Principle of Natural Self-Organization;* Springer-Verlag: Berlin, Germany, 1979.
53. Kauffman, S. A. *The Origins of Order: Self-Organization and Selection in Evolution;* Oxford University Press: New York, 1993.
54. Holland, H. D. *The Chemical Evolution of the Atmosphere and Ocean;* Princeton University Press: Princeton, NJ, 1984.
55. Holland, H. D. *Geochim. Cosmochim. Acta* **2002,** *66,* 3811–3826.
56. Kasting, J. F. *Science* **2001,** *293,* 819–820.
57. Fike, D. A.; Grotzinger, J. P.; Pratt, L. M.; Summons, R. E. *Nature* **2006,** *444,* 744–747.
58. Canfield, D. E.; Poulton, S. W.; Narbonne, G. M. *Science* **2007,** *315,* 92–95.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch001

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Chapter 2

The Origin of Biomolecules

H. James Cleaves II[1] and Antonio Lazcano[2]

[1]Geophysical Laboratory, Carnegie Institution of Washington, Washington, DC, 20015, USA
[2]Facultad de Ciencias, UNAM, Apartado Postal 70-407, Ciudad Universitaria, 04510 México DF, México

The ease of formation of amino acids, purines, pyrimidines, sugars and a wide variety of other organic compounds under plausible prebiotic conditions suggests that these molecules may have been present in the primitive terrestrial environment. It is likely that collisions of cometary nuclei with the primitive Earth, combined with the contribution from other extraterrestrial bodies such as meteorites and interplanetary dust particles, may have supplemented the primordial broth with extraterrestrial organic molecules. Submarine hydrothermal vents also likely played some role in prebiotic organic evolution. The primordial soup may have been a bewildering organic chemical wonderland, but it likely did not include all the compounds or molecular structures found in modern living systems. Regardless of their ultimate origin, simple organic compounds dissolved in the primitive oceans would need to be concentrated and polymerized by simple physicochemical mechanisms.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Introduction

How life started on Earth is not known. Although the processes that led to it remain elusive, most explanations suggest that the first forms of life were the evolutionary outcome of a complex mixture of organic compounds of abiotic origin; i.e., the discussion of the origin of life is necessarily a discussion of organic chemistry. Not surprisingly, some of our modern ideas on the origin of life have developed in tandem with discoveries in organic and biochemistry. This chapter will attempt to summarize experimental findings regarding the synthesis of organic building blocks that may have been important for the origin of life on Earth.

In 1828 Friedrich Wöhler demonstrated that heating ammonium cyanate would lead to urea, a result that represented the first synthesis of an organic compound from inorganic starting materials. A new era in chemical research had begun: in 1850 Adolph Strecker synthesized alanine in the laboratory from acetaldehyde, ammonia and hydrogen cyanide. This was followed by Butlerov's demonstration that the treatment of formaldehyde with alkaline catalysts leads to the synthesis of sugars. Since until the 1920's it was generally assumed that that the first living beings had been autotrophs, the abiotic formation of these organic compounds was not considered a necessary prerequisite for the origin of life. These syntheses were also not conceived of as prebiotic laboratory simulations, but rather as attempts to understand the autotrophic mechanisms of nitrogen assimilation and CO_2 fixation in green plants.

The situation changed with the proposal of an heterotrophic origin of life made in 1924 by A. I. Oparin (*1*), a young Russian biochemist. Oparin was convinced that it was impossible to reconcile his Darwinian beliefs in a gradual evolution of complexity with the commonly held suggestion that life had emerged already endowed with an autotrophic metabolism. He reasoned that since heterotrophic anaerobes were metabolically simpler than autotrophs, the former would necessarily have evolved first. Based on the simplicity of fermentative metabolism, Oparin (*1*) suggested that the first organisms must have been heterotrophic bacteria that could not make their own food but consumed organic material present in the primitive milieu.

Five years later John B.S. Haldane (*2*) independently published a similar hypothesis, which explains why such views are often credited to both scientists. Oparin's ideas were further elaborated in a more extensive book published in 1936 in Russian and two years later translated into English (*3*). In this new book, which is a major classic in evolutionary analysis, Oparin revised his original proposal, leading to the assumption of a highly reducing milieu in which iron carbides of geological origin would react with steam to form hydrocarbons. Their oxidation would yield alcohols, ketones, aldehydes, etc., that would then react with ammonia to form amines, amides and ammonium salts. The resulting protein-like compounds would form a hot dilute soup, which would aggregate to form colloids or coacervates, from which the first heterotrophic microbes evolved. Oparin did not address in his 1938 book the origin of nucleic acids, because at the time their role in genetic processes was not yet suspected (*3*).

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

For Oparin (*3*), highly reducing atmospheres corresponded to mixtures of CH_4, NH_3, and H_2O with or without added H_2. The atmosphere of Jupiter contains these chemical species, with H_2 in large excess over CH_4. Oparin's proposal of a primordial reducing atmosphere was a brilliant inference from the then fledgling knowledge of solar atomic abundances and planetary atmospheres. The benchmark contributions of Oparin's 1938 book include the hypothesis that heterotrophs and anaerobic fermentation were primordial, which led him to refine the idea of the proposal of a reducing atmosphere that could allow the prebiotic synthesis and accumulation of organic compounds. These ideas played a major role in shaping the views of Harold Clayton Urey, an avid experimentalist with a wide range of scientific interests that was interested in the composition of the early atmosphere based on then popular ideas of solar system formation. In 1952 Urey published *The Planets, their Origin and Development* (*4*), which delineated his ideas of the formation of the solar system, a formative framework into which most origin of life theories are now firmly fixed, albeit in slightly modified fashion.

However, not everybody accepted these ideas. In 1951 Rubey proposed an outgassing model based on an early core differentiation and assumed the early atmosphere would have been reminiscent of modern volcanic gases (*5*). In his model Rubey estimated that a CH_4 atmosphere could not have persisted for much more than 10^5 to 10^8 years due to photolysis (*5*). The Urey/Oparin atmospheric (CH_4, NH_3, H_2O) models are thus based on astrophysical and cosmochemical models, while Rubey's CO_2, N_2, H_2O model is based on extrapolation of the geological record. Although this early theoretical work has had a great influence on subsequent research, modern thinking on the origin and evolution of the chemical elements, the solar system, the Earth, and its atmosphere and oceans has not been shaped largely with the origin of life as a driving force. On the contrary, current origin of life theories have been modified to fit contemporary models in geo- and cosmochemistry.

The Primitive Terrestrial Atmosphere

The origin of life can be constrained into a relatively short period of the Earth's history. On the upper end, the age of the solar system has been determined to be approximately 4.65 billion years from isotopic data from terrestrial minerals, lunar samples and meteorites, and the Earth-moon system is estimated to be approximately 4.5 billion years old. The early age limit for the origin of life on Earth may also be constrained by the lunar cratering record, which suggests that the flux of large asteroids impacting the early Earth's surface until ~ 3.9 billion years ago was sufficient to boil the terrestrial oceans and sterilize the planet. On the more recent end, there is putative isotopic evidence for biological activity from ~3.8 billion years ago (living systems tend to incorporate the lighter isotope of carbon, ^{12}C preferentially over ^{13}C during carbon fixation due to metabolic kinetic isotope effects). There is more definitive fossil evidence from ~ 3.5 billion years ago in the form small organic inclusions in cherts morphologically similar to cyanobacteria, as well as stromatolitic assemblages (layered mats reminiscent of the layered deposits created by modern microorganismal communities). Thus the time window for the origin of life appears to be between ~3.9 billion and 3.8 billion years ago. Life would have had to originate in a relatively short period, and the synthesis and accumulation of the organic compounds for this event must have preceded it in an even shorter time period (*6*). The synthesis and survival of organic biomonomers on the primitive Earth would have depended on the prevailing environmental conditions. Unfortunately, to a large degree these conditions are poorly constrained by geological evidence.

The temperature at which the planets accreted is important for understanding the early Earth's atmosphere, which in turn is essential for the possibility of terrestrial prebiotic organic synthesis. This depends on the rate of accretion. If the planet accreted slowly, more of the primitive atmosphere may have been derived from planetesimals, likely reminiscent of the reducing chemistry of the early solar nebula, and this could have been retained. If it accreted rapidly, an idea which is currently favored, the original atmosphere would have been lost and the primitive atmosphere would have been the result of outgassing of retained volatiles and subsequent extraterrestrial delivery of volatiles. CH_4, CO_2, CO, NH_3, H_2O and H_2 are the most abundant molecular gas species in the solar system, and this was likely true on the early Earth as well, although their relative proportions is unknown. It remains contentious whether the Earth’s water was released via volcanic exhalation of water associated with minerals accreted during planetary formation or whether it was accreted from comets and other extraterrestrial bodies during planet formation.

It seems unlikely that the Earth kept much of its earliest atmosphere during early accretion, thus the primordial atmosphere would have been derived from outgassing of the planet's interior, which is thought to have occurred at temperatures between 300 to 1500° C. Modern volcanoes emit a wide range of gas mixtures, most of which is CO_2 and SO_2, rather than CH_4 and H_2S. However, it seems likely that most of the gases released today are from the reactions of reworked crustal material and water, and do not represent components of the Earth's deep interior. Thus modern volcanic gases may tell us little about the early Earth's atmosphere.

It is likely that the oxidation state of the early mantle governed the distribution of reducing gases released during outgassing. Holland (*7*) proposed a multistage model based on the Earth being formed through cold homogeneous accretion in which the Earth's atmosphere went through two stages, an early reduced stage before complete differentiation of the mantle, and a later neutral/oxidized stage after differentiation. During the first stage, the redox state of the mantle was governed by the Fe°/Fe^{2+} redox pair, or iron-wustite buffer. The atmosphere in this stage would be composed of H_2O, H_2, CO and N_2, with approximately a $0.27 - 2.7 \times 10^{-5}$ atm of H_2. Once Fe° had segregated into the core, the redox state of magmas would have been controlled by the Fe^{2+}/Fe^{3+} pair, or fayalite-magnetite-quartz buffer.

Little is agreed upon about the composition of the early atmosphere other than that it almost certainly contained very little free oxygen. Free oxygen can be produced by the photo-dissociation of water, but today this occurs rather slowly, and it seems likely that the steady state would have been low early in the Earth's history due to reaction with reduced metals in the crust and oceans such

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

as Fe^{2+}. Although it is generally accepted that free oxygen was absent from the early Archean Earth, there is no agreement on the composition of the primitive atmosphere; opinions vary from strongly reducing (CH_4 +N_2/NH_3 + H_2O, or CO_2 + H_2 + N_2) to neutral (CO_2 + N_2 + H_2O).

If a reducing atmosphere was required for terrestrial prebiotic organic synthesis, the crucial question becomes the source of H_2. Miller and Orgel (*8*) have estimated the pH_2 as 10^{-4} to 10^{-2} atm. Molecular hydrogen could have been supplied to the primitive atmosphere by several sources. First if there had been extensive weathering of Fe^{2+} bearing rocks which had not been equilibrated with the mantle, followed by photo-oxidation in water (*9*), although this reaction may also have been equilibrated during volcanic outgassing. The oxidation sate of the atmosphere is important for the production of HCN, which is essential for the synthesis of amino acids and purine nucleobases, as well as cyanoacetylene for pyrimidine nucleobase synthesis. In CH_4/N_2 atmospheres HCN is produced abundantly (*10, 11*), but in CO_2/N_2 atmospheres most of the N atoms produced by splitting N_2 recombine with O atoms to form NO_x species.

Urey's early assumptions as to the constitution of the primordial atmosphere lead to the landmark 1953 Miller-Urey experiment that succeeded in producing copious amounts of organic compounds including many important in modern biochemistry (*12*). Yields of intermediates as a function of the oxidation state of the gases involved have been investigated and it has been shown that reduced gas mixtures are generally much more conducive to organic synthesis than oxidizing or neutral gas mixtures. This appears to be because of the likelihood of reaction terminating O radical collisions where the partial pressure of O containing species is high. Even mildly reducing gas mixtures produce copious amounts of organic compounds, and it seems likely that energy was the not the limiting factor (*11*).

Prebiotic Synthesis of Amino Acids

The first successful prebiotic amino acid synthesis was carried out with an electric discharge (Figure 1) and a strongly reducing model atmosphere of CH_4, NH_3, H_2O, and H_2 (*12*). The result of this experiment was a large yield of racemic amino acids, together with hydroxy acids, short aliphatic acids, and urea. One of the surprising results of this experiment was that the products were not a random mixture of organic compounds, but rather a relatively small number of compounds were produced in substantial yield. Moreover, with a few exceptions, the compounds were of biochemical significance.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

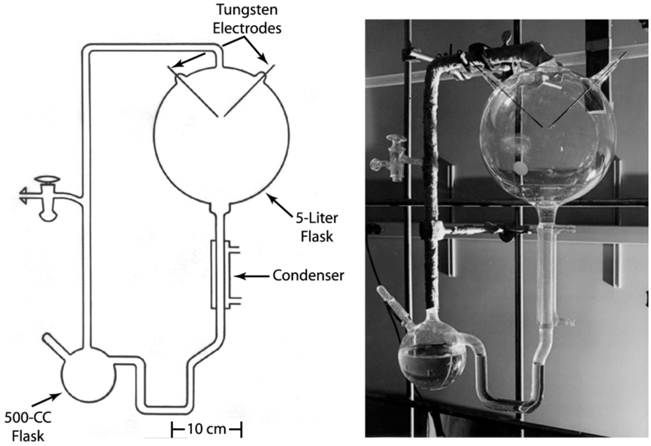


Figure 1. The Miller-Urey apparatus. (Reproduced from reference 14. Copyright 1957 Elsevier.)

The presence of large quantities of hydrogen cyanide, aldehydes and ketones in the water flask (Figure 2), which were clearly derived from the methane, ammonia, and hydrogen originally included in the apparatus (Figure 2), showed that the amino acids were not formed directly in the electric discharge, but were the outcome of a Strecker-like synthesis that involved aqueous phase reactions of reactive intermediates (*13*). Both amino- and hydroxy acids can be synthesized at high dilutions of HCN and aldehydes in a simulated primitive ocean (*14*). The reaction rates depend on temperature, pH, HCN, NH_3, and aldehyde concentrations, but are rapid on geologic time scales: the half-lives for the hydrolysis of the intermediate products in the reactions, amino- and hydroxy-nitriles, are less than a thousand years at 0° C (*15*). An additional example of a rapid prebiotic synthesis is that of amino acids on the Murchison meteorite (which will be discussed later), where it apparently took in a period of time as fast as 10^4 years (*16*). These results suggest that if the prebiotic environment was reducing, then the synthesis of the building blocks of life was efficient and did not constitute the limiting step in the emergence of life.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

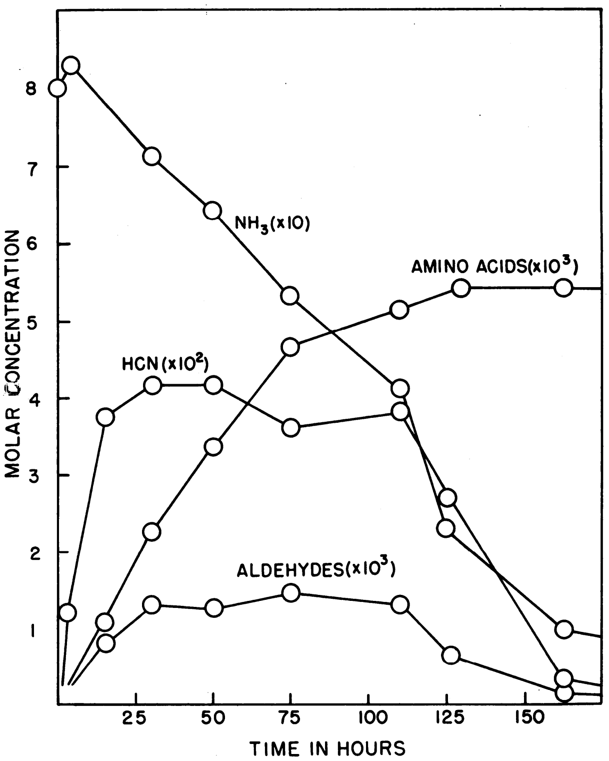


Figure 2. Variations in the concentrations of ammonia, aldehydes and hydrogen cyanide in the Miller-Urey experiment.

The Strecker synthesis of amino acids requires the presence of ammonia (NH_3) in the prebiotic environment. As gaseous ammonia is rapidly decomposed by ultraviolet light (*17*), during early Archean times the absence of a significant ozone layer would have limited ammonia's atmospheric concentration (Figure 3). However, since ammonia is very soluble in water, if the primitive oceans had a pH ~8, then dissolved NH_4^+ (the pK_a of NH_3 is ~ 9.2) in equilibrium with dissolved NH_3 would have been available. Since NH_4^+ is similar in size to K^+ and thus easily enters the same exchange sites on clays, NH_4^+ concentrations were probably no higher than 0.01 M. A more realistic atmosphere for the primitive Earth may be a mixture of CH_4 with N_2 with traces of NH_3. There is experimental evidence that this mixture of gases is quite effective with electric discharges in producing amino acids (*16*). Such an atmosphere, however, would nevertheless be strongly reducing. Alternatively, amino acids can be synthesized from the reaction of urea, HCN, and an aldehyde or a ketone, via the Bucherer-Bergs synthesis (Figure 4). This reaction pathway may have been significant if little ammonia were available.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

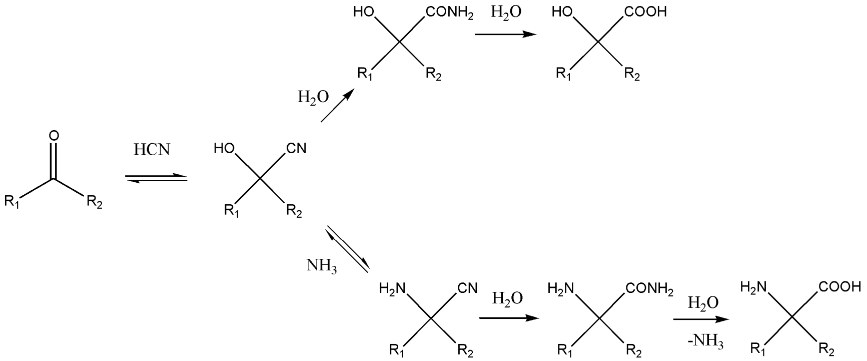


Figure 3. The Strecker and cyanohydrin mechanisms for the formation of amino- and hydroxy-acids from ammonia, cyanide, aldehydes and ketones.

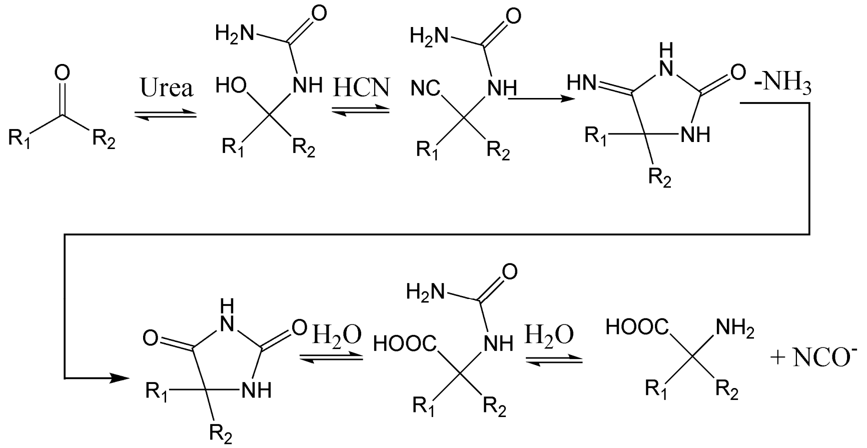


Figure 4. The Bucherer-Bergs reaction mechanism.

A wide variety of direct sources of energy must have been available on the primitive Earth. It is likely that in the prebiotic environment solar radiation, and not atmospheric electricity, was the major source of energy reaching the Earth's surface. However, it is unlikely that any single one of the energy sources can account for all organic compound syntheses. The importance of a given energy source in prebiotic evolution is determined by the product of the energy available and its efficiency in generating organic compounds. Given our current ignorance of the prebiotic environment, it is impossible to make absolute assessments of the relative significance of these different energy sources. For instance, neither the pyrolysis (800° to 1200° C) of a CH_4/NH_3 mixture or the action of ultraviolet light acting on a strongly reducing atmosphere give good yields of amino acids. However, the pyrolysis of methane, ethane, and other hydrocarbons gives good yields of phenylacetylene, which upon hydration yields phenylacetaldehyde. The later could then participate in a Strecker synthesis and act as a precursor to the amino acids phenylalanine and tyrosine in the prebiotic ocean (*15*).

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Experiments suggest that electric discharges were the most important source of hydrogen cyanide, which is an important intermediate in prebiotic synthesis. However photochemical mechanisms could also have been significant (*18*). In addition to its central role in the formation of amino nitriles during the Strecker synthesis, HCN polymers have been shown to be a source of amino acids. Ferris et al. (*19*) demonstrated that in addition to urea, guanidine, and oxalic acid, hydrolysis of HCN polymers produces glycine, alanine, aspartic acid, and α-aminoisobutyric acid. Modern organisms construct their proteins from ~ 20 universal amino acids which are almost exclusively of the L enantiomer. The amino acids produced by prebiotic syntheses are racemic. It is unlikely that all of the modern amino acids were present in the primitive environment, and it is unknown which if any may have been important for the origin of life.

Acrolein would have been produced in fairly high yield from the reaction of acetaldehyde with HCHO (*20*), which has several very robust atmospheric syntheses. Acrolein can be converted into several of the biological amino acids via reaction with various plausible prebiotic compounds (*21*) (Figure 5).

There has been less experimental work with gas mixtures containing CO and CO_2 as carbon sources instead of CH_4, although CO-dominated atmospheres could not have existed except transiently. Spark discharge experiments using CH_4, CO, or CO_2 as a carbon source with various amounts of H_2 have shown that methane is the best source of amino acids, but CO and CO_2 are almost as good if a high H_2/C ratio is used (Figure 6). Without added hydrogen, however, the amino acid yields are often quite low, especially when CO_2 is the sole carbon source. Recent results, however, suggest that amino acid yields from neutral atmospheres may be much better than previously thought (*22*). For example, it was found that buffering the solution pH near neutral and addition of oxidation inhibitors increased the amino acid yields from $CO_2/N_2/H_2O$ electric discharge reactions by several hundred fold.

The amino acid diversity produced in CH_4 experiments are similar to those reported by Miller (*12*). With CO and CO_2, however, glycine was the predominant amino acid, with little else besides alanine produced (*15*).

The implication of these results is that CH_4 is the best carbon source for abiotic synthesis. Although glycine was essentially the only amino acid produced in spark discharge experiments with CO and CO_2, as the primitive ocean matured the reaction between glycine, H_2CO, and HCN could have lead to the formation of other amino acids such as alanine, aspartic acid, and serine. Such simple mixtures may have lacked the chemical diversity required for prebiotic evolution and the origin of the first life forms. However, since it is not known which amino acids were required for the emergence of life, we can say only that CO and CO_2 are less favorable than CH_4 for prebiotic amino acid synthesis, but that amino acids produced from CO and CO_2 may have been adequate. The spark discharge yields of amino acids, HCN, and aldehydes are about the same using CH_4, $H_2/CO > 1$, or $H_2/CO_2 > 2$. However, it is not clear how such high molecular hydrogen-to-carbon ratios for the last two reaction mixtures could have been maintained in the prebiotic atmosphere.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

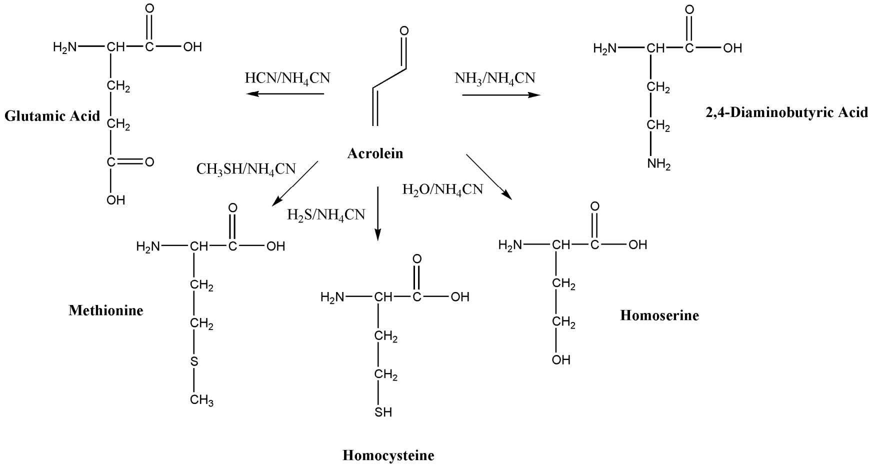

Glutamic Acid
HCN/NH₄CN
NH₃/NH₄CN
2,4-Diaminobutyric Acid
Acrolein
CH₃SH/NH₄CN
H₂O/NH₄CN
H₂S/NH₄CN
Methionine
Homoserine
Homocysteine

Figure 5. Acrolein as a precursor in the prebiotic synthesis of biological amino acids.

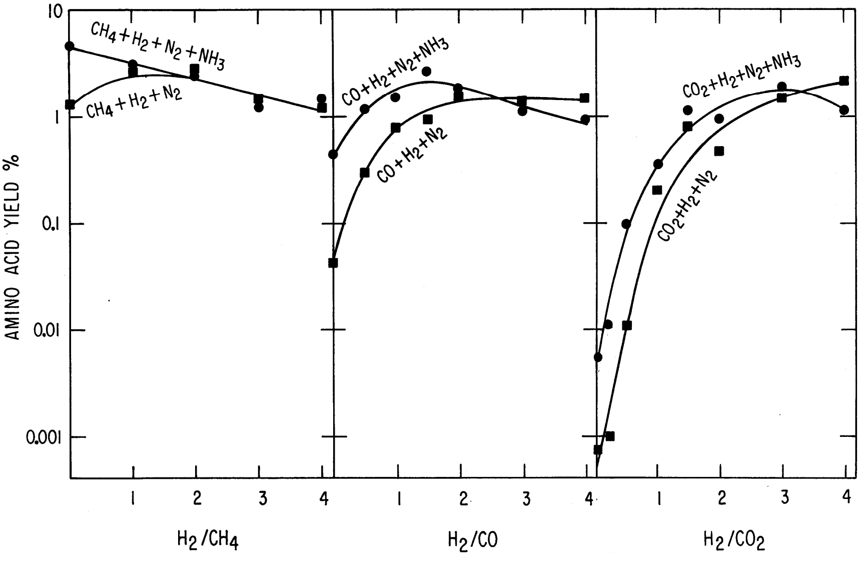


Figure 6. Amino acid yields based on initial carbon. In the experiments summarized here, the partial pressures of N_2, CO and CO_2 were 100 torr. The flask contained 100 ml of water but no NH_3. Experiments were conducted at room temperature, and the spark generator was operated continuously for two days. (Reproduced from reference 121. Copyright 1983 Springer.)

Synthesis of Nucleic Acid Components

In all known living organisms genetic information is stored in DNA, which is composed of repeating units of deoxyribonucleotides, which in turn is

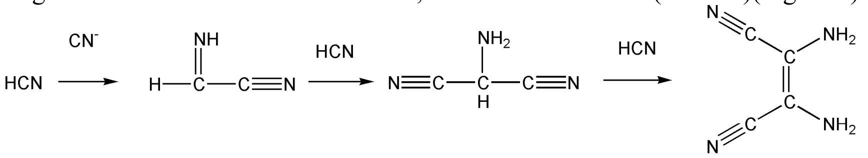
Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

transcribed into linear polymers of RNA, which are composed of repeating polymers of ribonucleotides. The difference between these is the use of deoxyribose in DNA and ribose in RNA, and uracil in RNA and thymine in DNA. Proposals for the existence of an "RNA World", a period during early biological evolution when biological systems used RNA both as a catalyst and an informational macromolecule, suggest that ribose may have been among the oldest carbohydrates to be employed by living beings. Although it is now doubted that RNA was formed directly from the primordial soup, experimental simulations of the prebiotic environment have shown the ease of formation of purines, pyrimidines and sugars. As shown long ago by Butlerov (*23*) together with the other sugars that are produced by the condensation of formaldehyde under alkaline conditions, it is also one of the organic compounds synthesized in the laboratory under prebiotically relevant conditions. Non-enzymatic phosphorylation reactions are also discussed below, although they may be unlikely prebiotic processess.

Purines

The first evidence that the components of nucleic acids may have been synthesized abiotically was provided in 1960 when Oró (*24*) reported the abiotic formation of adenine, which may be considered a pentamer of HCN ($C_5H_5N_5$) from aqueous NH_4CN. Oró found concentrated solutions of ammonium cyanide, which were refluxed for a few days, produced adenine in up to 0.5% yield along with 4-aminoimidazole-5 carboxamide and an intractable polymer (*24, 25*). The polymer also yields amino acids, urea, guanidine, cyanamide and cyanogen. It is surprising that a synthesis requiring at least 5 steps should produce such high yields of adenine. The mechanism of synthesis has since been studied in some detail. The initial step is the dimerization of HCN followed by further reaction to give HCN trimer and HCN tetramer, diaminomaleonitrile (DAMN)(Figure 7).



Aminomalononitrile
(HCN Trimer)

Diaminomaleononitrile
(HCN Tetramer)

Figure 7. The mechanism of formation of DAMN from HCN.

As demonstrated by Ferris and Orgel (*26*), a two-photon photochemical rearrangement of diaminomaleonitrile proceeds readily with high yield in sunlight to amino imidazole carbonitrile (AICN) (Figure 8). Other purines including guanine, hypoxanthine, xanthine, and diaminopurine could have been produced in the primitive environment by variations of the adenine synthesis using aminoimidazole carbonitrile and aminoimidazole carboxamide (*27, 28*) with other small molecule intermediates generated from HCN. Further reactions of AICN with other small molecules generated in polymerizing HCN have been shown to produce purines (Figure 9). These mechanisms, however, are likely an oversimplification. In dilute aqueous solutions adenine synthesis may also involve the formation and rearrangement of other precursors such as 2-cyano and 8-cyano adenine (*29*).

The steady state concentrations of HCN would have depended on the pH and temperature of the early oceans and the input rate of HCN from atmospheric synthesis. Assuming favorable production rates, Miyakawa *et al.* (*30*) estimated steady state concentrations of HCN of 2 x 10^{-6} M at pH 8 and 0°C in the primitive oceans. At 100° C and pH 8 the steady state concentration was estimated as 7 x 10^{-13} M. HCN hydrolyzes to formamide which then hydrolyzes to formic acid and ammonia. It has been estimated that oligomerization and hydrolysis compete at approximately 10^{-2} M concentrations of HCN at pH 9 (*31*), although it has been shown that adenine is still produced from solutions as dilute as 10^{-3} M (*32*). If the concentration of HCN were as low as estimated, it is possible that HCN tetramer formation may have occurred on the primitive Earth in eutectic solutions of HCN-H_2O, which may have existed in the polar regions of an Earth of the present average temperature. High yields of the HCN tetramer have been reported by cooling dilute cyanide solutions to temperatures between –10° C and –30° C for a few months (*31*). Production of adenine by HCN polymerization is accelerated by the presence of formaldehyde and other aldehydes, which could have also been available in the prebiotic environment (*29*).

Reexamination of the polymerization of concentrated NH_4CN solutions has shown that in addition to adenine, guanine is also produced at both -80° C and -20° C (*33*). It is probable that most of the guanine obtained from the polymerization of NH_4CN is the product of 2, 6-diaminopurine hydrolysis, which reacts readily with water to give guanine and isoguanine. Adenine, guanine, and a simple set of amino acids dominated by glycine have also been detected in dilute solutions of NH_4CN which were kept frozen for 25 years at -20° and -78° C, as well as in the aqueous products of spark discharge experiments from a reducing experiment frozen for 5 years at -20° C (*34*).

A wide variety of other modified purines may also have been available on the early Earth. The list includes isoguanine, which is a hydrolytic product of diaminopurine (*33*), and several modified purines which may have resulted from reactions of adenine or guanine with alkylamines under the concentrated conditions of a drying pond (*35*) including N^6-methyladenine, 1-methyl-adenine, N^6,N^6-dimethyladenine, 1-methylhypoxanthine, 1-methylguanine, and N^2-methylguanine.

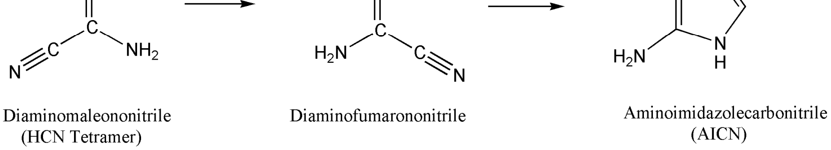


Figure 8. The synthesis of aminoimidazolecarbonitrile (AICN) via photoisomerization of DAMN.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

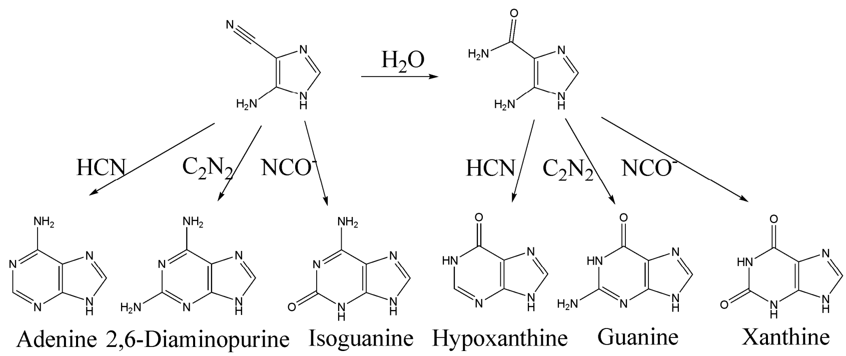


Figure 9. Production of purines from the reaction of AICN with other small molecules.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Pyrimidines

The prebiotic synthesis of pyrimidines has also been investigated extensively. The first synthesis investigated was that of uracil from malic acid and urea (*36*). The abiotic synthesis of cytosine in an aqueous phase from cyanoacetylene (HCCCN) and cyanate (NCO^-) was later described (*37, 38*). Cyanoacetylene is abundantly produced by the action of spark discharges on CH_4:N_2 mixtures, and cyanate is produced from cyanogen (NCCN) or from the decomposition of urea. Cyanoacetylene is also a Strecker synthesis precursor to aspartic acid. However, the high concentrations of cyanate (>0.1 M) required in this reaction are unrealistic, since cyanate is rapidly hydrolyzed to CO_2 and NH_3. Urea itself is fairly stable, depending on the concentrations of NCO^- and NH_3. Later, it was found that cyanoacetaldehyde (the hydration product of cyanoacetylene) and urea react to form cytosine and uracil. This was extended to a high yield synthesis that postulated drying lagoon conditions. The reaction of uracil with formaldehyde and formate gives thymine in good yield (*39*). Thymine is also synthesized from the UV catalyzed dehydrogenation of dihydrothymine, produced from the reaction of β-aminoisobutryic acid with urea (*40*).

The reaction of dilute aqueous cyanoacetaldehyde (produced from the hydrolysis of cyanoacetylene) with urea produces no detectable levels of cytosine (*41*). However, when the same non-volatile compounds are concentrated by evaporation in laboratory models of “drying lagoon” conditions simulating primitive pools on drying beaches on the early Earth, large yields of cytosine (> 50%) are obtained (*42*). These results suggest a facile mechanism for the accumulation of pyrimidines in the prebiotic environment (Figure 10). Evaporating cyanoacetaldehyde with guanidine produces 2, 4-diaminopyrimidine (*41*) in high yield (*43*), which then hydrolyses to uracil and small amounts of cytosine. Uracil (albeit in low yields), as well as its biosynthetic precursor orotic acid, were also identified among the hydrolytic products of hydrogen cyanide polymer (*19, 44*).

Modified pyrimidines may have also been present in the primitive Earth. These include dihydrouracil, which is formed from NCO^- and β-alanine (*45*), and others like 2, 4-diaminopyrimidine, 2-thiocytosine (*43*), and 5-substituted uracils, formed via reaction of uracil with formaldehyde, whose functional side groups may have played an important role in the early evolution of catalysis prior to the origin of proteins, and which are efficiently formed under plausible prebiotic conditions (*46*).

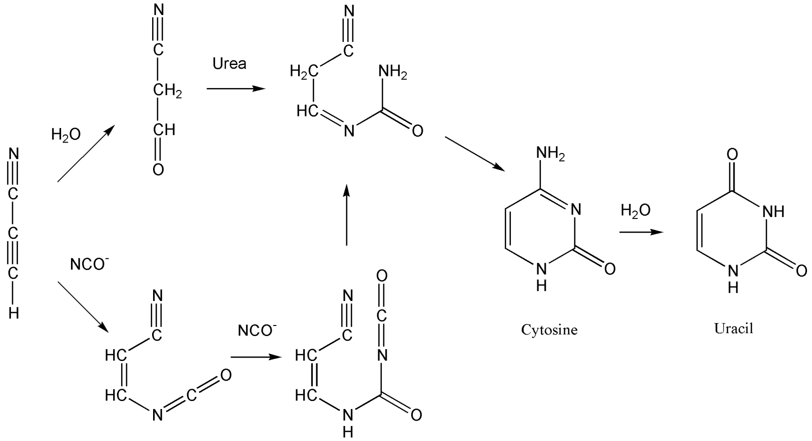


Figure 10. Two possible mechanisms for the prebiotic synthesis of pyrimidines from the reaction of cyanoacetaldehyde with urea concentrated by laboratory models of "drying lagoon".

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Carbohydrates

Most biological sugars have the empirical formula $(CH_2O)_n$, a point that was underscored by Butlerov's 1861 discovery of the formose reaction that showed that a diverse mixture of the sugars of biological relevance could be formed by the reaction of HCHO under basic conditions (*23*). However, the Butlerov synthesis is complex and incompletely understood. It depends on the presence of suitable inorganic catalysts, with calcium hydroxide ($Ca(OH)_2$) or calcium carbonate ($CaCO_3$) being the most completely investigated. In the absence of basic catalysts, little or no sugar is obtained. Clays such as kaolin serve to catalyze the formation of sugars, including ribose, in small yields from dilute (0.01 M) solutions of formaldehyde (*47-49*). This reaction has been extensively investigated with regard to catalysis and several interesting phenomena have been observed. For instance, the reaction is catalyzed by glycolaldehyde, acetaldehyde and various organic catalysts (*50*).

The Butlerov synthesis is autocatalytic and proceeds through glycoaldehyde, glyceraldehyde, and dihydroxyacetone, four carbon sugars, and five carbon sugars to give finally hexoses, including biologically important carbohydrates such as glucose and fructose. The reaction may proceed as shown in Figure 11. The reaction produces a complex mixture of sugars including various 3, 4, 5, 6 and 7-membered carbohydrates, including all isomers (for the addition of each CH_2O unit, both isomers are produced) and generally is not particularly selective, although methods of overcoming this have been investigated. Of special geochemical interest is the observation that borate can strongly skew the product mixture in favor of certain sugars, including ribose (*51, 52*).

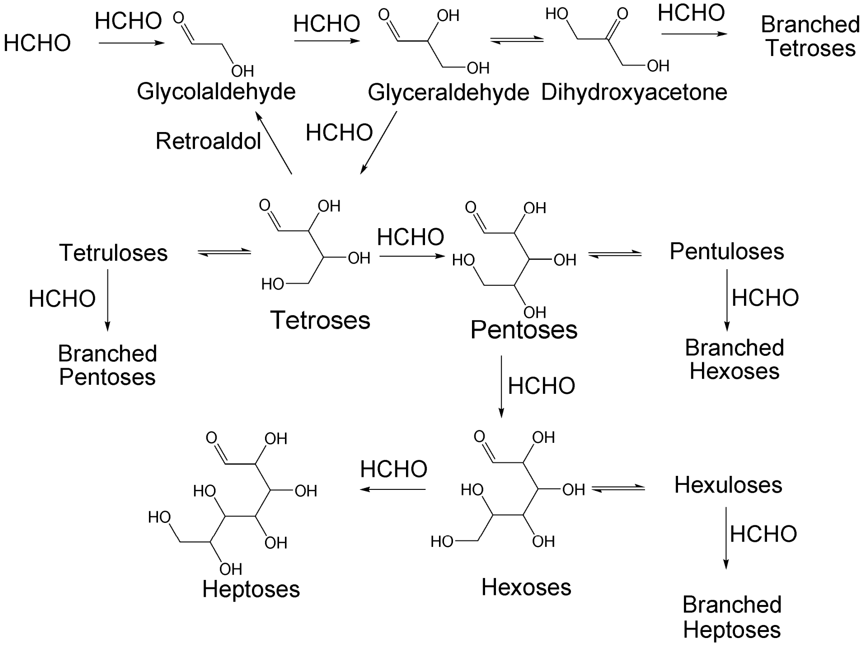

HCHO
HCHO
Glycolaldehyde
HCHO
Glyceraldehyde
Dihydroxyacetone
HCHO
Branched Tetroses
Retroaldol
HCHO
Tetruloses
HCHO
Branched Pentoses
Tetroses
HCHO
Pentoses
Pentuloses
HCHO
Branched Hexoses
HCHO
Heptoses
HCHO
Hexoses
Hexuloses
HCHO
Branched Heptoses

Figure 11. A simplified scheme of the formose reaction.

Inclusion of acetaldehyde in the reaction may lead to the formation of 2-deoxyribose (*53*). However, the reaction tends to stop when the formaldehyde has been consumed and ends with the production of higher C4-C7 sugars that can form cyclic acetals and ketals. The reaction produces all of the epimers and isomers of the small C2-C6 sugars, some of the C7 ones, and various dendroaldoses and dendroketoses, as well as small molecules such as glycerol and pentaerythritol.

Both L- and D-ribose occur in this complex mixture, but are not particularly abundant. Since all carbohydrates have somewhat similar chemical properties, it is difficult to envision simple mechanisms that could lead to the enrichment of ribose from this mixture, or how the relative yield of ribose required for the formation of RNA could be enhanced. However, the recognition that the biosynthesis of sugars leads not to the formation of free carbohydrates but of sugar phosphates, lead Albert Eschenmoser and his associates to show that under slightly basic conditions the condensation of glycoaldehyde-2-phosphate in the presence of formaldehyde considerable selectivity exist in the synthesis of ribose-2,4-diphosphate (*54*). This reaction has also been shown to take place under neutral conditions and low concentrations in the presence of minerals (*55*), and is particularly attractive given the properties of pyranosyl-RNA (p-RNA), a 2’,4’-linked nucleic acid analogue whose backbone includes the six-member pyranose form of ribose-2,4-diphosphate (*56*).

There are several obstacles to the relevance of the formose reaction as a source of sugars on the primitive Earth. One is that the Butlerov synthesis gives

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

a wide variety of straight chain- and branched sugars. More than 40 different sugars have been identified in one reaction mixture (*57*). Another problem is that the conditions of synthesis are also conducive to the degradation of sugars (*48*). Sugars undergo various diagenetic reactions on short geological time scales that are seemingly prohibitive to the accumulation of significant amounts on the primitive Earth. At pH 7, the half-life for decomposition of ribose is 73 minutes at 100° C, and 44 years at 0° C (*58*). The same is true of most other sugars, including ribose-2, 4-diphosphate. Finally, the concentrations of HCHO required appear to be prebiotically implausible, although the true limits of the synthesis have not been determined.

There are a number of possible ways to stabilize sugars; the most interesting one is to attach the sugar to a purine or pyrimidine, i.e., by converting the carbohydrate to a glycoside, but the synthesis of nucleosides is difficult under plausible prebiotic conditions. It has therefore been suggested that ribonucleotides could not have been the first components of prebiotic informational macromolecules (*59*). This has led to propositions of a number of possible substitutes for ribose in nucleic acid analogues, in what has been dubbed the "pre-RNA World" (*60*).

There is an additional paradox in prebiotic chemistry. When aqueous solutions of HCN and HCHO are mixed, the predominant product is glycolonitrile (*61*), which seems to preclude the formation of sugars and nucleic acid bases in the same location (*62*). Nevertheless both sugar derivatives and nucleic acid bases have been found in the Murchison meteorite (*63, 64*) and it seems likely that the chemistry which produced the compounds found in the Murchison meteorite was from aqueous reactions such as the Strecker synthesis. This suggests that the synthesis of sugars, amino acids and purines from HCHO and HCN can take place under certain combinations of NH_3, HCN, aldehydes and ketones and pH, or the different compound classes were produced under different regimes in different locations.

Lipids and Membrane-Forming Compounds

Amphiphilic molecules of sufficient chain length may spontaneously assemble into micelles and vesicles. Cell membranes are almost universally composed of phosphate esters of fatty acid glycerides. Fatty acids are biosynthesized today by multifunctional enzymes or enzyme complexes. As all life we know of is comprised of cells, these compounds seem crucial. Eukaryotic and bacterial cell membranes are composed of largely straight chain fatty acid acyl glycerols while those of the archaea are often composed of terpenoid glycerol ethers. Either type may have been the primordial lipid component of primitive cells. Most prebiotic simulations fail to generate large amounts of fatty acids, with the exception of simulated hydrothermal vent reactions, which arguably use unreasonably high concentrations of reactants (*65*). Heating glycerol with fatty acids and urea has produced acylglycerols (*66*). A prebiotic synthesis of long-chain isoprenoids has been suggested by Ourisson based on the Prins reaction of formaldehyde with isobutene (*67*). The Murchison meteorite contains small amounts of higher straight chain fatty acids, some of

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

which may be contamination (*68*). Amphiphilic components have been positively identified in the Murchison meteorite (*69*), although the yields of these molecules are poor in typical spark discharge experiments (*70*). There may have been other, more significant prebiotic sources of these molecules, however (*71, 72*).

Were There Prebiotic Nucleosides And Nucleotides?

There is considerable evidence for the existence of the so-called RNA world, a period during early biological evolution when RNA molecules played the role of both catalysts and genetic molecule (*73*). This has led to considerable debates on the prebiotic plausibility of ribonucleotides and their polymers. Much research has been carried out on the abiotic synthesis of nucleosides and nucleotides, although many consider that they do not correspond to truly prebiotic processes. Although few researchers still consider this idea plausible for the origin of life, it is possible that an RNA world existed as an intermediary stage in the development of life once a simpler self-replicating system had arisen.

Perhaps the most promising nucleoside syntheses start from purines and pure D-ribose in drying reactions, which simulate conditions that might occur in an evaporating basin. Using hypoxanthine and a mixture of salts reminiscent of those found in seawater, up to 8% of β-D-inosine is formed, along with the α-isomer. Adenine and guanine gave lower yields, and in both cases a mixture of α- and β-isomers was obtained (*74*). Pyrimidine nucleosides have proven to be much more difficult to synthesize. Direct heating of ribose and uracil or cytosine has thus far failed to produce uridine or cytidine. Pyrimidine nucleoside syntheses have been demonstrated which start from ribose, cyanamide, and cyanoacetylene, however α-D-cytidine is the major product (*75*). This can be photo-anomerized to β-D-cytidine in low yield, however the converse reaction also occurs. Sutherland and coworkers (*76*) demonstrated a more inventive approach, showing that cytidine-3'-phosphate can be prepared from arabinose-3-phosphate, cyanamide and cyanoacetylene in a one-pot reaction. The conditions may be somewhat forced, and the prebiotic source of arabinose-3-phosphate is unclear. Nevertheless the possibility remains that more creative methods of preparing the pyrimidine nucleosides may be possible.

Prebiotic phosphorylation of nucleosides is also somewhat problematic. Although there is some evidence that condensed phosphates are emitted in volcanic fumaroles (*77*), it has been suggested that condensed phosphates are not likely to be prebiotically abundant materials (*78*). Heating orthophosphate at relatively low temperatures in the presence of ammonia results in a high yield of condensed phosphates (*79*). Additionally, trimetaphosphate (TMP) has been shown to be an active phosphorylating agent for various prebiological molecules including amino acids and nucleosides (*80, 81*).

Early attempts to produce nucleotides used organic condensing reagents such as cyanamide, cyanate or dicyanamide (*82*). Such reactions were generally inefficient due to the competition of the alcohol groups of the nucleosides with water in an aqueous environment. Nucleosides can be phosphorylated with

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

acidic phosphates such as NaH_2PO_4 when dry heated (*83*). The reactions are catalyzed by urea and other amides, particularly if ammonium is included in the reaction. Nucleosides can also be phosphorylated in high yield by heating ammonium phosphate with urea at moderate temperatures, as might occur in a drying basin (*84*). For example, heating uridine with urea and ammonium phosphate gave yields of phosphorylated nucleosides as high as 70%. In the case of purine nucleosides, however there is also considerable glycosidic cleavage due to the acidic microenvironment created. This is another problem with the RNA world, that the synthesis of purine nucleosides is somewhat robust, but nucleotide formation may be difficult, while nucleotide formation from pyrimidine nucleosides is robust, but nucleoside formation appears to be difficult.

Common calcium phosphate minerals such as apatite are themselves reasonable phosphorylating reagents. Yields as high as 20% of nucleotides were achieved by heating nucleosides with apatite, urea and ammonium phosphate (*84*). Heating ammonium phosphates with urea leads to a mixture of high molecular weight polyphosphates (*79*). Although these are not especially good phosphorylating reagents under prebiotic conditions, they tend to degrade, especially in the presence of divalent cations at high temperatures, to cyclic phosphates such as trimetaphosphate, which have been shown to be promising phosphorylating reagents (*81*). Cis-diols, such as ribosides, react readily with trimetaphosphate under alkaline conditions to yield cyclic phosphates, and the reaction proceeds under neutral conditions in the presence of magnesium cation (*77*).

Alternatively, the difficulties with prebiotic ribose synthesis and nucleoside formation have led some to speculate that perhaps a simpler genetic molecule with a more robust prebiotic synthesis preceded RNA. Some propose substituting other sugars besides ribose (*85, 86*). When formed into sugar phosphate polymers, these also often form stable base-paired structures with both RNA/DNA and themselves (*87-89*), opening the possibility of genetic takeover from a precursor polymer to RNA/DNA. These molecules would likely suffer from the same drawbacks as RNA, such as the difficulty of selective sugar synthesis, sugar instability and the difficulty of nucleoside formation. Recently it has been demonstrated based on the work of Joyce *et al.* (*60*), and the chemistry proposed by Nelsestuen (*90*) and Tohidi and Orgel (*91*), that backbones based on acyclic nucleoside analogues may be more easily obtained under reasonable prebiotic conditions, for example, by the reaction of nucleobases with acrolein during mixed formose reactions (*92*). This remains a largely unexplored area of research.

More exotic alternatives to nucleosides have been proposed based on the peptide nucleic acid (PNA) analogues of Nielsen and coworkers (*93*). Miller and coworkers (*94*) were able to demonstrate the facile prebiotic synthesis of all of the components of PNA under the same chemical conditions required for the synthesis of the purines or pyrimidines. Nevertheless there are many alternative structures, which have not yet been investigated.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Hydrothermal Vents and the Origin Of Life

The discovery of hydrothermal vents at mid-ocean ridges and the appreciation of their significance in the element balance of the hydrosphere represents a major development in oceanography (*95*). Since the process of hydrothermal circulation probably began early in the Earth's history, it is likely that vents were present in the Archaean oceans. Large amounts of ocean water now pass through the vents, with the whole ocean going through them every 10 million years (*96*). This flow was probably greater during the early history of the Earth, since the heat flow from the planet's interior was greater during that period. The topic has received a great deal of attention, partly because of doubts regarding the oxidation state of the early atmosphere.

Following the first report of the vents' existence a detailed hypothesis suggesting a hydrothermal emergence of life was published (*97*), in which it was suggested that amino acids and other organic compounds are produced during passage through the temperature gradient of the 350° C vent waters to the 0° C ocean waters. Polymerization of the organic compounds thus formed, followed by their self-organization was also proposed to take place in this environment, leading to the first forms of life. At first glance, submarine hydrothermal springs appear to be ideally suited for creating life, given the geological plausibility of a hot early Earth. Vents exist along the active tectonic areas of the Earth, and at least in some of them catalytic clays and minerals interact with an aqueous reducing environment rich in H_2, H_2S, CO, CO_2, CH_4, and NH_3. Unfortunately it is difficult to corroborate these speculations with the composition of the effluents of modern vents, as much of the organic material released from modern sources is simply heated biological material, and it is difficult to separate the biotic from the abiotic components of these reactions.

Many proposals of a hydrothermal origin of life also suggest an autotrophic origin of life. So far, the most articulate autotrophic hypothesis stems from the work of Wächtershäuser (*98, 99*), who has argued that life began with the appearance of an autocatalytic, two-dimensional chemolithtrophic metabolic system based on the formation of the highly insoluble mineral pyrite (FeS_2). The FeS/H_2S combination is a strong reducing agent, and has been shown to provide an efficient source of electrons for the reduction of organic compounds under mild conditions. The scenario proposed by Wächtershäuser (*98, 99*) fits well with the environmental conditions found at deep-sea hydrothermal vents, where H_2S, CO_2, and CO are quite abundant. However, the FeS/H_2S system does not reduce CO_2 to amino acids, purines, or pyrimidines, although there is more than adequate free energy to do so (*100*). However, pyrite formation can produce molecular hydrogen, and reduce nitrate to ammonia, and ethyne to ethene (*101*). More recent experiments have shown that the activation of amino acids with carbon monoxide and (Ni, Fe)S can lead to peptide-bond formation (*102*).

In general organic compounds are decomposed rather than created at hydrothermal vent temperatures, although of course temperature gradients exist. As has been shown by Sowerby and coworkers (*103*) concentration on mineral surfaces would tend to concentrate any organics created at hydrothermal vents in cooler zones. The presence of reduced metals and the high temperatures of hydrothermal vents have also led to suggestions that reactions similar to those in

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Fischer-Trospch-Type (FTT) syntheses may be common under such regimes. It is unclear to what extent this is true, as typical FTT catalysts are easily poisoned by water and sulfide, although some of the likely environmental catalysts such as magnetite may be immune to such poisoning (*104*).

The Stability of Biomolecules at High Temperatures

Proposals of a thermophilic origin of life are not new, but few have been able to overcome the fact that high temperatures are also destructive to most bioorganic compounds. As noted above, some progress has been made in synthesizing small molecules under conditions simulating hydrothermal vents, but most biomolecules tend to decompose at high temperatures. Most biological molecules have half-lives to hydrolysis on the order of minutes to seconds at the very high temperatures associated with on-axis hydrothermal vents, and are still rather unstable at the somewhat lower temperatures of off-axis vent environments. Ribose and other sugars are very thermolabile compounds (*58*), but pyrimidines and purines, and many amino acids are nearly as labile. At 100° C the half-life for deamination of cytosine is 21 days, and 204 days for adenine (*59, 105*). Some amino acids are stable, e.g., alanine has a half-life for decarboxylation of approximately 1.9×10^4 years at 100° C, but serine decarboxylates to ethanolamine with a half-life of 320 days (*106*). White (*107*) measured the decomposition of various compounds at 250° C and pH 7 and found half-lives of amino acids from 7.5 s to 278 min, half-lives for peptide bonds from < 1 min to 11.8 min, half lives for glycoside cleavage in nucleosides from < 1 s to 1.3 min, decomposition of nucleobases from 15 – 57 min, and half-lives for phosphate esters from 2.3 to 420 minutes. It should be borne in mind that the half-lives for polymers would be even shorter as there are so many more potential breakage points in a polymer. Thus while the vents may serve as synthesis sites for simpler compounds such as acetate or more refractory organic compounds such as fatty acids, it is unlikely they played a major role in synthesizing most biochemicals or their polymers.

Submarine vents do not seem to presently synthesize organic compounds more complex than very simple hydrocarbons such as CH_4 and ethane, more likely they decompose them over short time spans ranging from seconds to a few hours. Thus, the origin of life in the vents appears improbable. This does not imply that the hydrothermal springs were a negligible factor on the primitive Earth. If the mineral assemblages were sufficiently reducing, the rocks near the vents may have been a source of atmospheric CH_4 or H_2. As stated earlier, the concentrations of biomolecules which could have accumulated on the primitive Earth is governed largely by the rates of production and the rates of destruction. Submarine hydrothermal vents would have been important in the destruction rather than in the synthesis of organic compounds, thus fixing the upper limit for the organic compound concentrations in the primitive oceans. Although we do not presently know which compounds were essential for the origin of life it seems possible to preclude certain environments if even fairly simple organic compounds were involved (*108*).

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Extraterrestrial Contribution of Organic Compounds

Regardless of what the early Earth's atmosphere was like, the planet was undoubtedly bombarded then, as now, by extraterrestrial material such as meteorites. The presence of organic compounds in meteorites was recognized since the mid-19th century, when Berzelius analyzed the Aläis meteorite. Today the presence of a complex array of extraterrestrial organic molecules in meteorites, comets, interplanetary dust and interstellar molecules is firmly established, and has lead some to propose them as sources of the prebiotic organic compounds necessary for the origin of life (*109-112*).

In spite of some early debates, the occurrence of organic materials in the solar system was confirmed in 1969 when a meteorite fell in Murchison Australia. This meteorite was seen to fall and was rapidly collected, thus minimizing field contamination, and analyzed in the laboratory. A host of organic compounds was found to be present which were indubitably of extraterrestrial origin. These organics strongly resemble those produced in laboratory syntheses under presumed prebiotic conditions.

The flux of extraterrestrial organics to the Earth has been estimated based on the lunar cratering record (*112*). Exogenous delivery could have made a significant contribution to the primitive Earth's organic inventory. The estimated flux of HCN equivalents compared with that produced in a reducing atmosphere is shown in Figure 12. Thus even if the early Earth’s atmosphere were oxidizing or neutral, significant amounts of prebiotic organic compounds resembling the types made via terrestrial atmospheric synthesis under reducing conditions could have been delivered to the Earth.

What was the relative percentage of prebiotic organic matter contributed by each source? Although there is a wide variety of interstellar molecules, including formaldehyde, hydrogen cyanide, acetaldehyde, cyanoacetylene and other prebiotic compounds, and their total amount in a nebular dust cloud is high, it is generally felt that they played at most a minor role in the origin of life. The major sources of exogenous compounds would appear to be comets and dust, with asteroids and meteorites being minor contributors. Asteroids would have impacted the Earth frequently in prebiotic times, but the amount of organic material brought in would have been small, even if the asteroid were a Murchison meteorite-type object. The same considerations would apply to carbonaceous chondrites, although the survival of the organics on impact would be much better than with asteroids. A significant percentage of meteoritic amino acids and nucleobases could survive the high temperatures associated with frictional heating during atmospheric entry, and become part of the primitive broth (*113*).

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

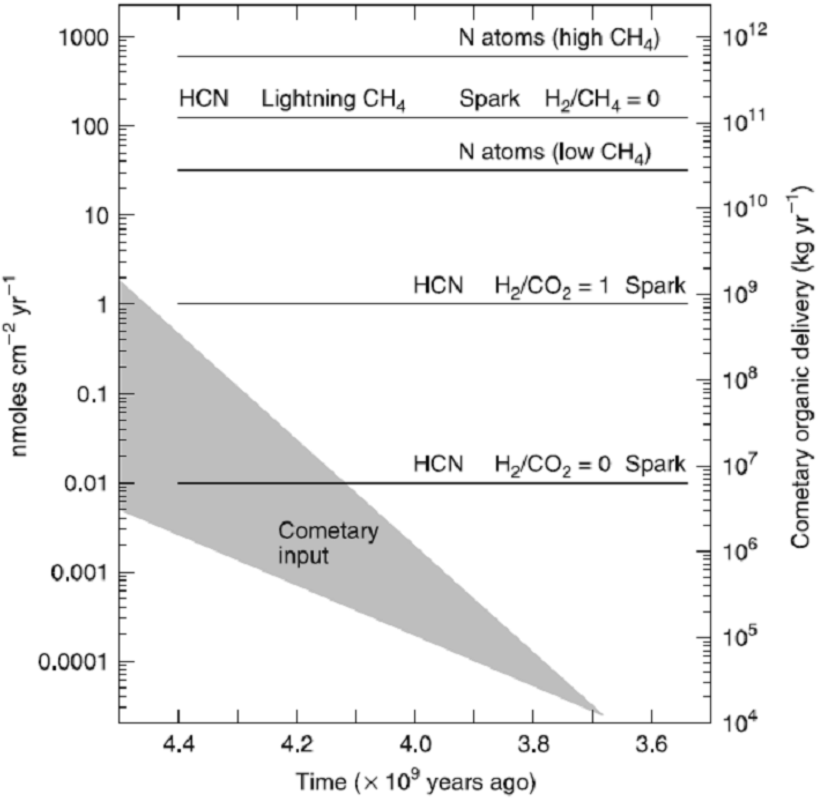


Figure 12. Fluxes of HCN to the surface of the early Earth estimated from endogenous syntheses compared with cometary sources vs. time. (Reproduced from reference 15. Copyright 1998 Cambridge University Press.)

Comets are one of the most promising sources of exogenous compounds (*109, 114*). Cometary nuclei contain about 80 % H_2O, 1 % HCN, and 1% H_2CO, as well as important amounts of CO_2 and CO. Thus, assuming that cometary nuclei have a 1 gm cm^{-3} density, a single 1 km-diameter comet would contain 2 x 10^{11} moles of HCN, or 40 nmoles cm^{-2} of the Earth's surface. This is comparable to the yearly production of HCN in a reducing atmosphere from electric discharges, and would be quite important if the Earth did not have a reducing atmosphere. This calculation assumes a complete survival of the HCN on impact. In fact, there is little understanding of what happens during the impact of such an object, but much of it would be heated to temperatures above 300° C which would decompose HCN and other cometary compounds. However, these highly reactive chemical species could then be used as precursors in the abiotic syntheses of biochemical monomers.

The input from interplanetary dust particles (IDPs) may have been even more important. The present infall is quite large (*115*), but on the primitive Earth it may have been greater by a factor of 100 to 1000. Unfortunately, the organic composition of IDPs is poorly known (*116*). The only individual molecules that have been detected are polycyclic aromatic hydrocarbons and α-

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

aminoisobutyric acid (*117, 118*). Organic polymers called tholins, which are produced by electric discharges, ionizing radiation, and ultraviolet light, could be a major component of IDPs. Although quite resilient, a few percent of amino acids are released from tholins on acid hydrolysis. A more promising role for the tholins would be as a source of prebiotic precursors such as HCN, cyanoacetylene, and aldehydes. On entry to the Earth's atmosphere, the dust particles would be heated and the tholins pyrolyzed, creating HCN and other molecules, which could then participate in prebiotic reactions (*119, 120*).

Summary

Given adequate expertise and experimental conditions, it is possible to synthesize almost any organic molecule in the laboratory under simulated prebiotic conditions. However, the fact that a number of molecular components of contemporary cells can be formed in the laboratory does not necessarily mean that they were essential for the origin of life, or that they were available in the prebiotic milieu. The primordial soup must have been a complex mixture, but it could not reasonably have included all the compounds found today in even the simplest cells. The basic tenet of the heterotrophic theory of the origin of life is that the origin and reproduction of the first living system depended primarily on abiotically synthesized organic molecules. As summarized here, there has been no shortage of discussion about how the formation of the primordial soup took place. It is unlikely that any single mechanism can account for the wide range of organic compounds that may have accumulated on the primitive Earth. The prebiotic soup was undoubtedly formed by contributions from endogenous atmospheric synthesis, deep-sea hydrothermal vent synthesis, and exogenous delivery from sources such as comets, meteorites and interplanetary dust. This eclectic view does not beg the issue of the relative significance of the different sources of organics; it simply recognizes the wide variety of potential sources of organic compounds, the raw material required for the emergence of life.

There are many mechanisms by which biochemical monomers can be synthesized under plausible prebiotic conditions. Not all prebiotic pathways are equally efficient, but the wide range of experimental conditions under which organic compounds can be synthesized demonstrates that prebiotic syntheses of the building blocks of life are robust, i.e., the abiotic reactions leading to them do not take place under a narrow range defined by highly selective reaction conditions, but rather under a wide variety of environmental settings. Although our ideas on the prebiotic synthesis of organic compounds are based largely on experiments in model systems, the robustness of this type of chemistry is supported by the occurrence of most of these biochemical compounds in the Murchison meteorite. It is therefore plausible that similar synthesis took place on the primitive Earth.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Acknowledgments

We are indebted to Dr. John Chalmers and Professor Jeffrey L. Bada for helpful comments.

References

1. Oparin, A. I. *Proiskhozhedenie Zhizni;* Mosckovskii Rabochii: Moscow, USSR, 1924. Reprinted and translated in J. D. Bernal *The Origin of Life;* Weidenfeld and Nicolson: London, 1967.
2. Haldane, J. *The Rationalist Annual* **1929,** 3–10.
3. Oparin, A. *The Origin of Life;* MacMillan: New York, 1938.
4. Urey, H. *The Planets, Their Origin and Development;* Yale University Press: New Haven, CT, 1952.
5. Rubey, W. *Geol. Soc. Am. Bull.* **1951,** *62,* 1111–1148.
6. Lazcano, A.; Miller, S. L. *J. Mol. Evol.* **1994,** *39,* 546–54.
7. Holland, H. *Petrologic Studies: A Volume to Honor A. F. Buddington;* Geological Society of America: New York, 1962; pp 447–477.
8. Miller, S.; Orgel, L. *The Origins of Life on the Earth;* Prentice-Hall: Englewood Cliffs, NJ, 1974.
9. Mauzerall, D. C. *Origins of Life Evol. Biosphere* **1990,** *20,* 293–302.
10. Chameides, W.; Walker, J. *Origins of Life* **1981,** *11,* 291–302.
11. Stribling, R.; Miller, S. *Origins of Life Evol. Biosphere* **1987,** *17,* 261–273.
12. Miller, S. *Science* **1953,** *117,* 528.
13. Miller, S. *J. Am. Chem. Soc.* **1955,** *77,* 2351–2361.
14. Miller, S. *Biochimica et Biophysica Acta* **1957,** *23,* 480–489.
15. Miller, S. In *The Molecular Origins of Life: Assembling the Pieces of the Puzzle;* Brack, A., Ed.; Cambridge University Press: Cambridge, UK, 1998; pp 59–85.
16. Peltzer, E.; Bada, J.; Schlesinger, G.; Miller, S. *Adv. Space Res.* **1984,** *4,* 69–74.
17. Kuhn, W.; Atreya, S. *Icarus* **1979,** *37,* 207–213.
18. Zahnle, K. *J. Geophys. Res.* **1986,** *91,* 2819–2834.
19. Ferris, J. P.; Joshi, P. D.; Edelson, E. H.; Lawless, J. G. *J. Mol. Evol.* **1978,** *11,* 293–311.
20. Cleaves, H. *Monatsh. Chem.* **2003,** *134,* 585–593.
21. Van Trump, J.; Miller, S. *Science* **1972,** *178,* 859–860.
22. Cleaves, H. J.; Chalmers, J. H.; Lazcano, A.; Miller, S. L.; Bada, J. L. *Origins of Life Evol. Biosphere* **2008,** *38,* 105–115.
23. Butlerov, A. *Ann. Chem.* **1861,** *120,* 29.
24. Oró, J. *Biochem. Biophys. Res. Commun.* **1960,** *2,* 407–412.
25. Oró, J.; Kimball, A. P. *Arch. Biochem. Biophys.* **1961,** *94,* 221–227.
26. Ferris, J.; Orgel, L. *J. Am. Chem. Soc.* **1966,** *88,* 1074.
27. Sanchez, R.; Ferris, J.; Orgel, L. *J. Mol. Biol.* **1967,** *30,* 223–253.
28. Sanchez, R.; Ferris, J.; Orgel, L. *J. Mol. Evol.* **1968,** *38,* 121–128.
29. Voet, A.; Schwartz, A. *Bioorg. Chem.* **1983,** *12,* 8–17.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

30. Miyakawa, S.; Cleaves, H.; Miller, S. L. *Origins of Life Evol. Biosphere* **2002,** *32,* 195–208.
31. Sanchez, R.; Ferris, J.; Orgel, L. *Science* **1966,** *153,* 72–73.
32. Miyakawa, S.; Cleaves, H.; Miller, S. *Origins of Life Evol. Biosphere* **2002,** *32,* 209–218.
33. Levy, M.; Miller, S.; Oró, J. *J. Mol. Evol.* **1999,** *49,* 165–168.
34. Levy, M.; Miller, S.; Brinton, K.; Bada, J. *Icarus* **2000,** *145,* 609–613.
35. Levy, M.; Miller, S. *J. Mol. Evol.* **1999,** *48,* 631–637.
36. Fox, S.; Harada, K. *Science* **1961,** *133,* 1923–1924.
37. Sanchez, R.; Ferris, J.; Orgel, L. *Science* **1966,** *154,* 784–785.
38. Ferris, J.; Sanchez, R.; Orgel, L. *J. Mol. Biol.* **1968,** *33,* 693–704.
39. Choughuley, A.; Subbaraman, A.; Kazi, Z.; Chadha, M. *BioSystems* **1977,** *9,* 73–80.
40. Schwartz, A.W.; Chittenden, G. *BioSystems* **1977,** *9,* 87–92.
41. Ferris, J.; Zamek, O.; Altbuch, A.; Freiman, H. *J. Mol. Evol.* **1974,** *3,* 301–309.
42. Robertson , M.; Miller, S. *Nature* **1995,** *375,* 772–774.
43. Robertson, M.; Levy, M.; Miller, S. L. *J. Mol. Evol.* **1996,** *43,* 543–550.
44. Voet, A.; Schwartz, A. *Origins of Life* **1982,** *12,* 45–49.
45. House, C.; Miller, S. *Biochem.* **1996,** *35,* 315–320.
46. Robertson, M.; Miller, S. *Science* **1995,** *268,* 702–705.
47. Gabel, N.; Ponnamperuma, C. *Nature* **1967,** *216,* 453–455.
48. Reid, C.; Orgel, L. *Nature* **1967,** *216,* 455.
49. Schwartz, A.; De Graaf, R. *J. Mol. Evol.* **1993,** *36,* 101–106.
50. Matsumoto, T.; Komiyama, M.; Inoue, S. *Chem. Lett.* **1980,** *7,* 839–842.
51. Prieur, B. E. *C R Acad Sci II*; **2001,** *4,* 667–670.
52. Ricardo, A.; Carrigan, M. A.; Olcott, A. N.; Benner, S. A. *Science* **2004,** *303,* 196.
53. Oró, J. In *The Origins of Prebiological Systems and of Their Molecular Matrices;* Fox, S.W., Ed.; Academic Press: New York, 1965; pp 137–162.
54. Muller, D.; Pisch, S.; Kittaka, A.; Wagner, E.; Wintner, C. E.; Eschenmoser, A. *Helv. Chim. Acta* **1990,** *73,* 1410–1468.
55. Pitsch, S.; Krishnamurthy, R.; Bolli, M.; Wendeborn, S.; Holzner, A.; Minton, M.; Leseur, C.; Schlönvogt, I.; Jaun, B.; Eschenmoser, A. *Helv. Chim. Acta* **1995,** *78,* 1621–1635.
56. Beier. M.; Reck, F.; Wagner, T.; Krishnamurthy, R.; Eschenmoser, A. *Science* **1999,** *283,* 699–703.
57. Decker, P.; Schweer, H.; Pohlmann, R. *J. Chrom.* **1982,** *225,* 281–291.
58. Larralde, R.; Robertson, M.; Miller, S. *Proc. Natl. Acad. Sci. USA* **1995,** *92,* 8158–8160.
59. Shapiro, R. *Origins Life Evol. Biosph.* **1995,** *25,* 83–98.
60. Joyce, G.; Schwartz, A.; Miller, S.; Orgel, L. *Proc. Natl. Acad. Sci. USA* **1987,** *84,* 4398–4402.
61. Schlesinger, G.; Miller, S. *J. Am. Chem. Soc.* **1973,** *95,* 3729–3735.
62. Arrhenius, T.; Arrhenius, G.; Paplawsky, W. *Origins of Life Evol. Biosphere* **1994,** *24,* 1–17.
63. Van der Velden, W.; Schwartz, A. *Geochim. Cosmochim. Acta* **1977,** *41,* 961–968.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

64. Cooper, G.; Kimmich, N.; Belisle, W.; Sarlnana, J.; Brabham, K.; Garrel, L. *Nature* **2001,** *414,* 879–883.
65. McCollum, T.; Ritter, G.; Simoneit, B. *Origins of Life Evol. Biosphere* **1999,** *29,* 153–166.
66. Hargreaves, W.; Mulvihill, S.; Deamer, D. *Nature* **1977,** *266,* 78–80.
67. Ourisson, G.; Nakatani, Y. *Chem. & Biol.* **1994,** *1,* 11–23.
68. Yuen, G.; Kvenvolden, K. *Nature* **1973,** *246,* 301–303.
69. Deamer, D. W. *Nature* **1985,** *317,* 792–794.
70. Allen, W.; Ponnamperuma, C. *Curr. Mod. Biol.* **1967,** *1,* 24–28.
71. Deamer, D.; Dworkin, J. P.; Sandford, S. A.; Bernstein, M. P.; Allamandola, L. J. *Astrobiology* **2002,** *2,* 371–381.
72. Monnard, P. A.; Deamer, D. W. *Anat. Rec.* **2002,** *268,* 196–207.
73. *The RNA World 2nd Ed.;* Gesteland, R.; Cech, T.; Atkins, J., Eds.; Cold Spring Harbor: NY, 1999.
74. Fuller, W.; Sanchez, R.; Orgel, L. *J. Mol. Evol.* **1972,** *1,* 249–257.
75. Sanchez, R.; Orgel, L. *J. Mol. Biol.* **1970,** *47,* 531–543.
76. Ingar, A.; Luke, R.; Hayter, B.; Sutherland, J. *ChemBioChem* **2003,** *4,* 504–507.
77. Yamagata, Y.; Watanabe, H.; Saitoh, M.; Namba, T. *Nature* **1991,** *352,* 516–519.
78. Keefe, A.; Miller, S. *J. Mol. Evol.* **1995,** *41,* 693–702.
79. Osterberg, R.; Orgel, L. *J. Mol. Evol.* **1972,** *1,* 241–248.
80. Rabinowitz, J.; Hampai, A. *Helv. Chim. Acta* **1978,** *61,* 1842–1847.
81. Schwartz, A. *J. Chem. Soc. D: Chem. Commun.* **1969,** *23,* 1393.
82. Lohrmann, R.; Orgel, L. *Science* **1968,** *161,* 64–66.
83. Beck, A.; Lohrmann, R.; Orgel, L. *Science* **1967,** *157,* 952.
84. Lohrmann, R.; Orgel, L. *Science* **1971,** *171,* 490–494.
85. Eschenmoser, A. *Science* **1999,** *284,* 2118–2124.
86. Eschenmoser, A. *Origins of Life Evol. Biosphere* **2004,** *34,* 277–306.
87. Schwartz, A.; Orgel, L. *Science* **1985**, *228,* 585–587.
88. Schneider, K.; Benner, S. *J. Am. Chem. Soc.* **1990,** *112,* 453–455.
89. Eschenmoser, A. *Origins of Life Evol. Biosphere* **2004,** *34,* 277–306.
90. Nelsestuen, G. *J. Mol. Evol.* **1980,** *15,* 59–72.
91. Tohidi, M.; Orgel, L. *J. Mol. Evol.* **1989,** *28,* 367–373.
92. Cleaves, H. *Astrobiol.* **2002,** *2,* 403–415.
93. Nielsen, P.; Egholm, M.; Berg, R.; Buchardt, O. *Science* **1991,** *254,* 1497–1500.
94. Nelson, K.; Levy, M.; Miller, S. *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 3868–3871.
95. Corliss, J.; Dymond, J.; Gordon, L.; Edmond, J.; von Herzen, R.; Ballard, R.; Green, K.; Williams, D.; Bainbridge, A.; Crane, K.; van Andel, T. *Science* **1979,** *203,* 1073–1083.
96. Edmond, J.; Von Damn, K.; McDuff, R.; Measures, C. I. *Nature* **1982,** *297,* 187–191.
97. Corliss, J.; Baross, J.; Hoffman, S. *Oceanologica Acta* **1981,** *4 Suppl.,* 59–69.
98. Wächtershäuser, G. *Microbiol. Rev.* **1988,** *52,* 452–484.
99. Wächtershäuser, G. *Prog. Biophys. Mol. Biol.* **1992,** *58,* 85–201.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

100. Keefe, A.; Newton, G.; Miller, S. *Nature* **1995,** *373,* 683–685.
101. Maden, B. *Trends Biochem. Sci.* **1995,** *20,* 337–341.
102. Huber, C.; Wächtershäuser, G. *Science* **1998,** *281,* 670–672.
103. Sowerby, S. J.; Morth, C. M.; Holm, N. G. *Astrobiology* **2001,** *1,* 481–487.
104. Holm, N.; Andersson, E. In *The Molecular Origins of Life: Assembling the Pieces of the Puzzle*; Brack, A., Ed.; Cambridge University Press: Cambridge, UK, 1998; pp 86–99.
105. Garrett, E.; Tsau, J. *J. Pharm. Sci.* **1972,** *61,* 1052–1061.
106. Vallentyne, J. *Geochim. Cosmochim. Acta* **1964,** *28,* 157–188.
107. White, R. *Nature* **1984,** *310,* 430–432.
108. Cleaves, H.; Chalmers, J. *Astrobiol.* **2004,** *4,* 1–9.
109. Oró, J. *Nature* **1961,** *190,* 442–443.
110. Anders, E. *Nature* **1989,** *342,* 255-257.
111. Chyba, C. *Nature* **1990,** *343,* 129–133.
112. Chyba, C.; Sagan, C. *Nature* **1992,** *355,* 125–132.
113. Glavin, D.; Bada, J. *Astrobiol.* **2001,** *1,* 259–269.
114. Oró, J.; Lazcano, A. In *Comets and the Origin and Evolution of Life;* Thomas, P. J.; Chyba, C. F.; McKay, C. P., Eds.; Springer: New York, 1997; pp 3–27.
115. Love, S.; Brownlee, D. *Science* **1993,** *262,* 550–553.
116. Maurette, M. In *The Molecular Origins of Life;* Brack, A., Ed.; Cambridge University Press: Cambridge, UK, 1998; pp 147–186.
117. Gibson, E. *J. Geophys. Res.* **1992,** *97,* 3865–3875.
118. Clemett, S.; Maechling, C.; Zare, R.; Swan, P.; Walker, R. *Science* **1993,** *262,* 721–725.
119. Mukhin, L.; Gerasimov, M.; Safonova, E. *Nature* **1989,** *340,* 46–48.
120. Chyba, C.; Thomas, P.; Brookshaw, L.; Sagan, C. *Science* **1990,** *249,* 366–373.
121. Schlesinger, G.; Miller, S. L. *J. Mol. Evol.* **1983,** *19,* 376–382.

Downloaded by COLUMBIA UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch002

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

Chapter 3

The Hydrothermal Source of Energy and Materials at the Origin of Life

Michael J. Russell[1] and Allan J. Hall[2]

[1]**Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109-8099, USA**
[2]**Department of Archaeology, University of Glasgow, G12 8QQ, Scotland**

The case is presented that chemosynthetic life emerged at the growing exterior of a submarine hydrothermal mound generated where effluent from an alkaline hydrothermal convection cell interfaced a mildly acidic ocean across a precipitated membrane over 4 billion years ago. At first a small portion of the carbonic ocean water entrained in the submarine convection cell feeding the mound was reduced at very slow rates to acetate and methane. Deposits of porous iron-nickel sulfides, freshly precipitated in the mound, further catalyzed these reactions. At the same time, amino acids and peptides, generated in this milieu from hydrothermal hydrogen, ammonia and carbon dioxide, were retained in the mound's pores. These peptides then sequestered the iron-nickel sulfide clusters to produce protoenzymes that further 'quickened' the acetate and methane reactions. RNAs generated in the system then took over a catalytic and coding role to help drive the emergence of the first microbes—the acetogens and the methanogens, which evolved into the first bacteria and archaea respectively.

"We are forced to work from energy upward into form, because, at the beginning, form is nothing, energy is everything." H. F. Osborn, 1917, p. xv

Osborn (*1*) was first to think carefully about the provision of energy to the origin and evolution of life. Following his dictum, the primary purpose of this paper will be to review the initial physical and chemical conditions on the early Earth and thus to outline the likely available energies and geochemical ingredients responsible for life's emergence—the opening chapter of evolution's story. Our model is predicated upon the view that geophysical and geochemical conditions on the early Earth must have been its cause and that the main geochemical potential that drove life into being was the disequilibrium between hydrothermal hydrogen and atmospheric carbon dioxide. However, because of kinetic and thermodynamic challenges, the first protobiotic reactions may have evolved from what are now considered C1 intermediates such as formate ($HCOO^-$) or methanol (CH_3OH), rather than carbon dioxide itself, and that a steep proton gradient was also involved.

Any attempt to understand the path from protobiotic reactions toward the first microbes must consider how anaerobic prokaryotes work. Yet to understand what the first genetically-regulated organisms were like we cannot just 'undo' a cell, look for the simplest bits and attempt to reproduce it by whatever means in the lab (*2*). For harking back from known microbiology stops us short at the RNA world—an 'event horizon' beyond which it is difficult to see. That is why we start from likely potential geochemical beginnings in order to surmise what went on before genetically-coded living cells emerged. In such a forward-looking perspective the RNA world provides a target for convergence with evolutionary microbiology as we puzzle out how and where chemical and electrochemical energy would have been trapped, and ultimately spent, in a journey leading us from aqueous geochemistry to the most likely early metabolic pathways and their attendant biochemistries, i.e., from inorganic beginnings to the first anaerobic autotrophs (*3-7*). While acknowledging the speculative nature of the hypothesis, for clarity we write as though the scenario and mechanisms we offer here is real.

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

Acidic and Alkaline Fluids

Because the early highly radioactive Earth had been battered with smaller planets and large meteorites it began as a molten sphere. This excess of heat was transferred from, and through differentiating concentric spheres *via* convective mass transfer, eventually to radiate to the cold sink of space. Effective though this transfer of physical energy was, the chemical disequilibrium between our rocky planet and its volatisphere (atmosphere and ocean) was thereby exacerbated (Figure 1). At high temperatures (≤1700°C), acidic and oxidized entities such as CO_2, CO, NO, SO_2 and P_4O_{10} were (and still are) exsolved from the convecting mantle to the ocean and atmosphere as magma rose to form volcanoes (*8-10*). These volatiles were stable because the Earth's mantle, having lost any native iron by early gravitation to the core, sequestered residual iron to the mineral perovskite in the ferric state and thus was in equilibrium with these volatiles (*11-13*). Even submarine springs at 400°C emit some carbon dioxide, though accompanied by hydrogen and some methane. However, at the lower temperature of around 120°C, more reduced entities (H_2, $HCOO^-$, NH_3, HS^-, CH_3S^-) were exhaled through convective cycling in alkaline hydrothermal systems—a result of hydration and oxidation (serpentinization) of the

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

magnesium-iron silicates comprising the oceanic crust (*14*) (Figure 2). Thus strong chemical and electrochemical disequilibria were focused where the volatiles and metals exhaled from high temperature volcanoes and hot springs and dissolved in the Hadean Ocean, were eventually juxtaposed to alkaline submarine spring waters at ≤120°C (Figure 1). And most of the chemical and electrochemical energy was first discharged at a submarine hydrothermal mound generated above such alkaline submarine springs. Alkaline springs only start to operate continuously when temperatures are poised low enough to allow brittle fractures to develop in the serpentinizing mafic crust, but high enough to establish a density contrast in the hydrothermal system sufficient to drive a convective updraft at a temperature inferred to be around 120ºC (*15-17*). These systems last at least 35,000 years (*18*)—in our opinion long enough for the first microbes to emerge from inorganic beginnings.

Alkaline Solutions in Steps Toward Carbon Metabolism

Our approach to the emergence of life requires an understanding of how energies were captured, stored and used in the generation of structure. Most chemical energy was available at a moderate temperature submarine hydrothermal mound due to the potential reaction between marine CO_2 and hydrothermal H_2. Thermodynamically we expect the reaction to produce methane and water (*19,20*):

$$CO_2 + 4H_2 \Rightarrow CH_4 + 2H_2O \qquad (1)$$

However, if H_2 and pH values of the fluid are relatively high and the temperature low, there is more of an overall thermodynamic drive in favor of acetate (*19*):

$$2CO_2 + 4H_2 \Rightarrow CH_3COO^- + H^+ + 2H_2O \qquad (2)$$

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

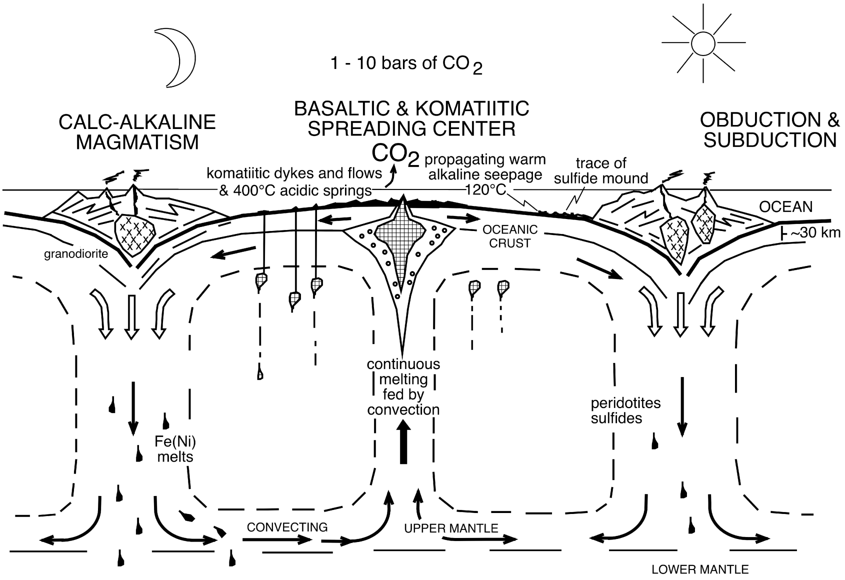

1 - 10 bars of CO2
CALC-ALKALINE MAGMATISM
BASALTIC & KOMATIITIC SPREADING CENTER
OBDUCTION & SUBDUCTION
CO2
komatiitic dykes and flows & 400°C acidic springs
propagating warm alkaline seepage 120°C
trace of sulfide mound
OCEAN
~30 km
OCEANIC CRUST
granodiorite
continuous melting fed by convection
peridotites sulfides
Fe(Ni) melts
CONVECTING
UPPER MANTLE
LOWER MANTLE

Figure 1. Initial conditions for the emergence of life. High temperature hydrothermal convection cells (discharging ~400 °C acidic spring waters) and moderate temperature hydrothermal convection cells (discharging ≤120 °C alkaline seepage waters) are both coupled to large-scale convection of the Earth's mantle. Note the gravitational loss of native iron-nickel melts, leaving the mantle oxidized enough to produce carbon dioxide from volcanoes. Thus, by 4 billion years or more ago, carbon dioxide dominated the atmosphere and ocean while hydrogen was delivered to the submarine mounds in the warm alkaline springs. Reactions between the hydrogen and carbon dioxide and other reduced and oxidized entities leading to the emergence of life at an alkaline seepage were mainly catalyzed by metal sulfide complexes. The metals in these complexes were ultimately supplied to the Hadean Ocean by the 400 °C acidic springs. Komatiitic and basaltic volcanism consists mainly of magnesium-iron silicates, and is centered over updrafts of convecting mantle and isolated pods of magma in the mantle. Calc-alkaline magma contains more calcium, sodium and potassium as its name implies. Calc-alkaline volcanism is generated at obduction (overthrust) and subduction (underthrust) zones. (Redrawn from reference 16. Copyright 2006 Geological Society of America.)

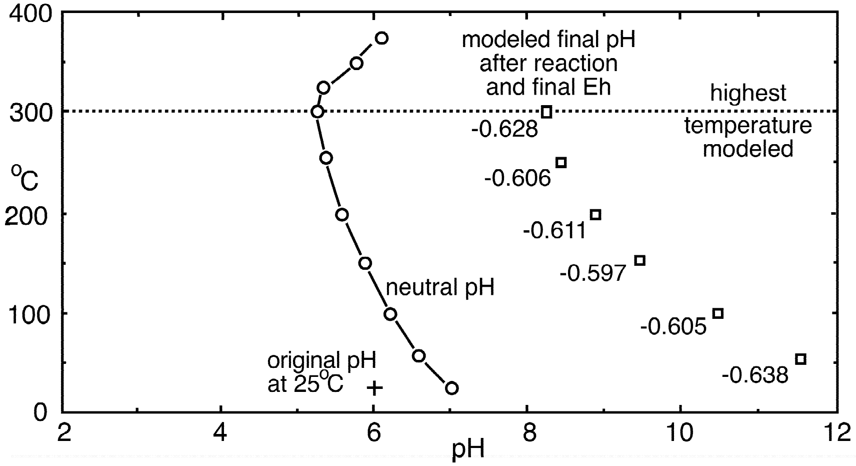


Figure 2. Plot of pH for the final computed aqueous hydrothermal solution produced on reacting model Hadean Ocean water with komatiite over a range of temperatures using the program SUPCRT. Neutral pH is drawn for reference. Note the reduced nature of the alkaline solution (Eh) recorded in volts. (Redrawn from reference 14 with permission from Springer.)

Both reactions are kinetically *and* thermodynamically (*21)* challenged, requiring energy and catalysis (and later, enzymes) for their resolution (Figure 3). The kinetic barrier is actually higher for methane synthesis because the addition of a hydride to a methyl group to make CH_4 faces more of an obstacle than does the addition of a further carbon-bearing radical to the same group to produce $H_3C\text{-}COO^-$ (*22*). Thus a route to acetate via a series of reactions, i.e., the acetyl coenzyme-A pathway, was almost certainly the first to develop (Figure 3) (*7,22*). While the acetogenic bacteria take this metabolic pathway, as we might expect, the methanoarchaea also use it (*23*). However, the ways in, and the ways out, of this pathway as taken by methanoarchaea are more complicated *(24)*. Nevertheless, the emergence of the methanoarchaea could not have been far behind in evolutionary terms (*7*).

Even the path to acetate is partly endergonic—there is not sufficient chemical energy available to reduce carbon dioxide beyond formate with any rapidity (*25*). This is where pyrophosphate comes into play. In prokaryotes four adenosine triphosphate (ATP) molecules are required to make the one 3-carbon triose-phosphate ($HCOH.CHOH.CH_2OPO(OH)_2$). Other core metabolisms need even more ATP: the reductive tricarboxylic acid cycle requires 5 ATPs per triosephosphate (*26*); the Calvin cycle requires 9; and the 3-hydroxyproprionate pathway needs 10 ATP (*26*). Thus from the principle of least work as well as its simplicity, the acetyl coenzyme-A pathway is likely to have been the basis of the first metabolism and the homoacetogens to have been the first microbes. Yet along this path to acetate only a single molecule of ATP is generated chemically from substrate (Figure 3)—insufficient to drive the pathway and subsequent biosynthesis. This energy shortfall has generally gone unrecognized in origin of life research. Yet the shortfall is probably the reason life exists at all. For if the intermediate entities in reaction 1 were all exergonic, then plain aqueous

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

geochemistry would see to the production of a methane atmosphere as carbon dioxide exsolving from magma and dissolved in the Hadean Ocean was continuously reduced.

So, how to make more pyrophosphate, or at least supply energy sufficient to overcome endergonic inhibition? The way extant life does it is to generate a protonmotive force by driving protons to the exterior of the cell and permitting a selective reintroduction through a highly evolved rotor–stator type ATPase (*27*). However, a compelling aspect of the alkaline scenario is that a biochemical invention of a system to drive protons to a cell's exterior was not required at life's emergence—adequate electrochemical (chemiosmotic) energy was already available to expedite the generation of pyrophosphate as a consequence of the pH gradient of around 5 units between the exterior acidic ocean and the interior of the hydrothermal mound saturated with alkaline solution. We can think of this energy as an ambient and inexhaustible protonmotive force (*16*); a force still required for the operation of all cells, though now produced by them (*28*).

However, even ATP is too complicated to be protobiotic. Instead we have assumed a surface reaction within the FeS membrane involving the formation and reformation of pyrophosphate for biosynthesis at a pH of ~ 8.3 as protons are driven down the pH gradient of around 5 units (*16,30,31*). It is notable that this equilibrium at high activity of phosphate intersects the stability field of FeS at high bisulfide activity (Figure 4). That pyrophosphates can be produced naturally is evidenced by the discoveries of a hydrated biphosphate (canaphite, $Ca_2Na_2PO_7.4H_2O$) and a triphosphate (kanonerovite, $MnNa_3P_3O_{10}.12H_2O$) (*32,33*).

It is this protonic energy that effects, via the generation of pyrophosphate, the stepwise reduction of carbon dioxide to a methyl group in the acetogens in the reverse 'Eastern branch' of the acetyl-CoA pathway (*36*) (Figures 3 and 5). Such a protonmotive force also contributes directly to the reduction of methyl thiol to methane in the methanogens. However, the generation of the $-COO^-$ entity (Figure 5) in the methanogens now requires a sodium gradient (*26*), though we presume the proton gradient did the job at the emergence of methanogenesis. So life's first vehicles, the homoacetogens and the methanoarchaea, required both chemical and electrochemical energy for their metabolisms—in modern parlance, they were 'hybrids'.

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

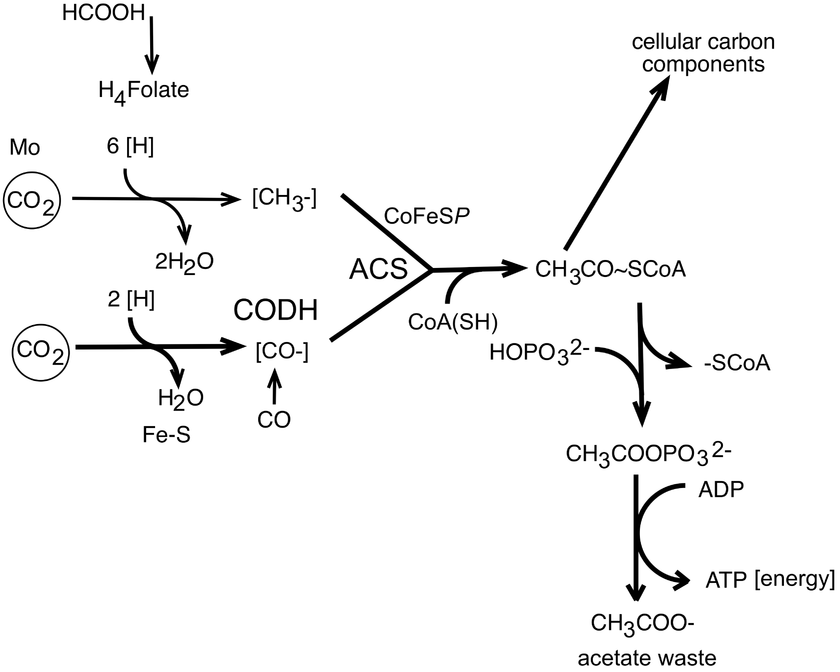


Figure 3. The acetyl coenzyme-A pathway as employed by the acetogens which produces both biochemical components and some energy in the form of adenosine triphosphate (ATP) (22). The acetogens are assumed here to have been the first microbes. Note that hydrothermal formic acid can feed into the "Eastern branch" (36) of the cycle via tetrahydrofolate (H_4Folate) in prokaryotes, ahead of carbon dioxide, by disproportionation to CO_2 and H_2. Carbon monoxide dehydrogenase/acetyl coenzyme-A synthase (CODH/ACS) is the bifunctional enzyme wherein one nickel is involved in the reduction of CO_2 (in the "Western branch") and another nickel is involved in synthesizing the acetyl group (29).

Early Biosynthesis

Activated acetate (as coA.SOC.CH_3 or originally as CH_3.SOC.CH_3) is generated in *both* the acetogenic (Figure 3) *and* the methanogenic biosynthetic pathways. And beginning with this activated C2 molecule, cellular carbon can be formed whereby each new useful molecular component would be derived from immediate simpler precursors rather than from condensations of the most simple. Pyruvate (CH_3.CO.COO^-), higher carboxylic acids and, with the addition of hydrothermal ammonia (*8*), amino acids and peptides, were early products. The bases were synthesized from various simple precursors much like they are today, but syntheses were unlikely to have involved cyanide (*7*).

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

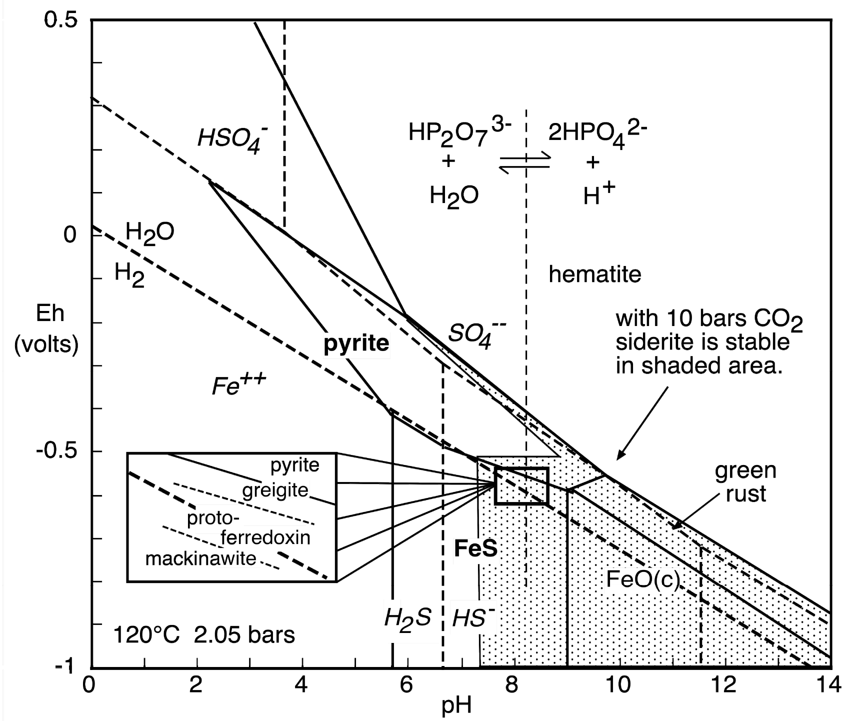


Figure 4. A Pourbaix diagram illustrating how the pH boundary of monophosphate/polyphosphate intersects the redox-zoned iron sulfide membrane; conditions pertaining to alkaline hydrothermal fluid as it enters the Hadean Ocean. During its active period the membrane probably consisted of a greigite (Fe_5NiS_8) exterior through to a mackinawite (FeS) interior with protoferredoxin occupying the mid zone as shown tentatively in the inset (16,30,34,66). The stabilities of siderite ($FeCO_3$), pyrite (FeS_2), green rust ($Fe_6(OH)_{12}CO_3.2H_2O$), and hematite (Fe_2O_3) are also shown. Diagram produced for activities of $H_2S(aq) = 10^{-3}$, and $Fe^{2+} = 10^{-6}$, using Geochemist's Workbench computing code (16,35). (Redrawn from reference 30. Copyright 1997 Geological Society of London.)

All these products would have been concentrated in pore space in the hydrothermal mound through a process of physical and chemical trapping followed by thermal diffusion (*39*). Eventually all the various geochemical processes leading to fully operating cells were conflated across scales of cubic kilometres, through metres and millimetres to microns, as a genetic system emerged, catalysts evolved to enzymes, and cofactors were invented (*7,30,40,41*).

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

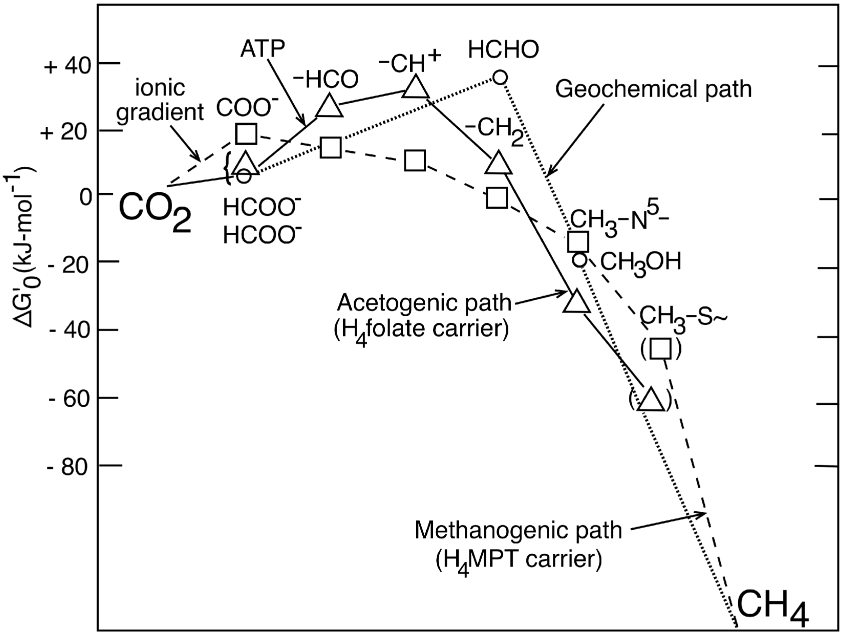


Figure 5. Free-energy profiles of acetogenic (triangles) and methanogenic (squares) reduction pathways (reverse of the 'Eastern branch' of the acetyl coenzyme-A pathway (36) compared to the geochemical pathway (circles) (37). We can think of the geochemical pathway as a chemical siphon while the much more rapid biochemical pathways are like chemical vortices. The homoacetogenic bacteria and the methanoarchaea are the probable root organisms of the two prokaryote domains, viz., the bacteria and the archaea (7). H_4*Folate is tetrahydrofolate and* H_4*MPT is tetrahydromethanopterin. Based on Maden (38).*

Metal, Phosphate, Sulfur, Formate, and Ammonia Sources in Acid and Alkaline Solutions

Here we explore the potential of volcanoes and acidic and alkaline solutions to provide all the entities required for life to emerge. Volcanoes provided the volatisphere with carbon dioxide substrate for life as well as pyrophosphate as P_4O_{10} (*42-43*). Magma-driven aqueous acidic hydrothermal springs operated at constructive and destructive plate margins. Their temperatures were regulated around 400ºC due to the buoyancy of water at its critical point (*44*). Depending on their depth, they exhaled up to 80 mmol/kg of ferrous iron into the early oceans (*45*). There was little to prevent a rapid build-up over time of ferrous ions in the Hadean Ocean to tens of millimoles per kilogram, though some may have been partitioned out in siderite ($FeCO_3$) during high temperature convective recharge (Figure 4) (*16*). Transition-metal to sulfide ratios in the exhaling hot acidic spring waters were high because the dearth of sulfate in the early oceans meant that, unlike the present, there was little source of sulfate sulfur for thermal reduction (*45*).

Iron from the ocean and sulfhydryl ion from some of the submarine alkaline springs supplied the constituents of the first membranes and iron-sulfur clusters required for electron transfer and hydrogenation (Figure 6). Nickel, which is important biochemically both as condensation and redox catalyst, would have tended to remain within the mafic and ultramafic crust as nickel-iron amalgams, as sulfides and in serpentinite silicates (*30*). Cobalt, vital to methyl transfer, is also very sparingly soluble in sulfidic solutions. However both nickel and cobalt have been found in cobaltian pentlandite ($\sim Co_{11}Ni_2FeCuS_{13}$) in hydrothermal precipitates at the Logatchev site on the Mid-Atlantic Ridge (*46*). Nickel and cobalt are also found commonly together, often as arsenides associated with mafic and ultramafic intrusives and ophiolites (*47,48*) suggesting they are more soluble in sulfide-poor solutions and thus would be more concentrated in the earliest oceans. Molybdenum and tungsten on the other hand, significant in the initial step of carbon dioxide fixation, would be contributed directly to the hydrothermal mound, dissolved as aqueous MoS_4^{2-} and WS_4^{2-} in the reduced alkaline fluids (*48-50*).

We consider the hydrothermal mound that developed above the alkaline submarine springs to be the hatchery of life (*30*). The growing outer surface of the mound acted as an inorganic catalytic membrane and compartment wall that inhibited mixing between the alkaline hydrothermal solution within, and the carbonic ocean without. In places it comprised iron sulfide dosed with other catalytic elements (e.g., Zn>Ni>Co>Mo>W). It formed as the alkaline spring waters met ferrous complexes dissolved in the acidulous Hadean Ocean. The unstable sulfides would have precipitated directly by condensation of soluble complexes around neutral pH ($pK_{FeS} = 10^{-5.7}$) (Figure 4). The sulfide complex $[2H_2O.FeSSFe.2H_2O]$ would have precipitated first as amorphous FeS—quickly crystallizing as nanocrystalline mackinawite on the margins of the hydrothermal mound (*30,51,52*). On burial and diagenesis of the mound the early mackinawite would be continually replaced as the rising alkaline fluid continued to invade the mound beneath, leaving ferrous hydroxide in its wake (Figure 6):

$$FeS + H_2O + OH^- \Leftrightarrow Fe(OH)_2 + HS^- \qquad (3)$$

In an approximate reversal of this same reaction, sulfhydryl ion (HS^-) would be responsible for further iron sulfide precipitation at the exterior of the growing mound. A more detailed equation in which water is the oxidizing agent and electron acceptor demonstrates how green-rust ($Fe_6(OH)_{12}CO_3.2H_2O$) and further hydrogen are produced (*53*):

$$6FeS + CO_3^{2-} + 12H_2O + 2OH^- \Rightarrow Fe^{II}_4Fe^{III}_2(OH)_{12}.CO_3.2H_2O + 6HS^- + 2H_2\uparrow \qquad (4)$$

Because of the metal-like (i.e., reduced and conducting) atomic structure of the [FeSSFe] rhombs comprising the mackinawite, and their harboring of nickel and cobalt atoms, the sulfide front or surface of the mound acts as catalyst and electron transfer agent during the emergence of life (*16,30*). Eventually such an atomic structure translates to the active center of hydrogenases, ferredoxins and

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

other iron(nickel) sulfide proteins (*16,52*). And because of the redox contrast across this crust there is zoning from mackinawite ([Fe>>Ni>Co]S) out to the somewhat more oxidized greigite ($NiFe_5S_8$), the second and last step along this precipitation pathway or side reaction (*53,54*).

The carbon dioxide and nitrogen required to build organic molecules were both carried from the atmosphere in ocean water that gravitated to depth in the downdraft limbs of the alkaline convection cells. While a proportion of the CO_2 carbonates the crust:

$$Mg_3Si_2O_5(OH)_4 + 3CO_2 \Rightarrow 3MgCO_3 + 2SiO_2 + 2H_2O \qquad (5)$$

much of the rest is reduced by hydrothermal hydrogen (see above) to formate or other forms of dissolved reduced carbon at neutral to alkaline conditions (*55,56*). McCollom and Seewald (*55,* p. 3640) have shown that the reaction:

$$HCO_3^- + H_2 \Rightarrow HCOO^- + H_2O \qquad (6)$$

"is rapid on geologic time scales at temperatures as low as 100°C." Alkaline fluids also favored other kinds of chemical interactions as life emerged (*16*). The formate is hydrogenated and sulfidized to methyl thiol in the mound, facilitated by the iron sulfides dosed with nickel, cobalt, as well as molybdenum and selenium. Nitrogen on the other hand, is reduced to ammonia during hydrothermal convection to the extent of around 40mM according to Shock (*8*, and see *57,58*), a reaction likely to continue in the sulfidic portions of the mound itself (*59*).

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

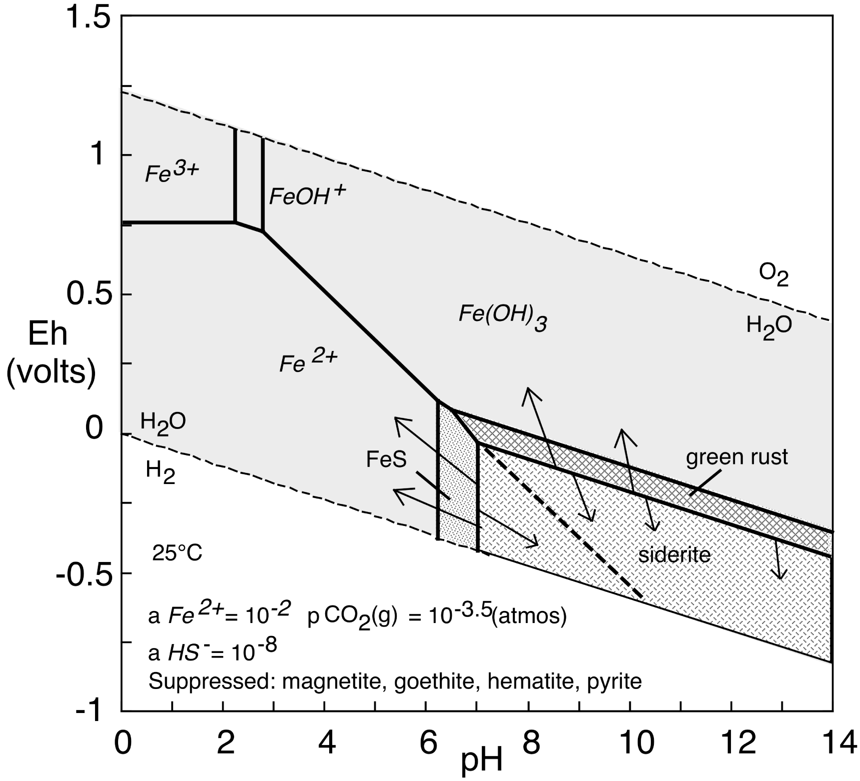


Figure 6. A Pourbaix diagram illustrating the stabilities of siderite ($FeCO_3$), mackinawite (as FeS) and green rust ($Fe^{II}_4Fe^{III}_2(OH)_{12}{\cdot}CO_3.2H_2O$) produced for activities of $H_2S(aq) = 10^{-3}$, and $Fe^{2+} = 10^{-6}$. At higher pCO_2 the siderite field would expand as indicated by arrows, at the expense of FeS. The FeS/green-rust boundary is projected as a dashed line to show its approximate position at very low pCO_2) in alkaline solution at depth in the mound. Release of HS– from mackinawite at depth (and from pyrrhotite [~Fe_9S_{10}] accumulations in the crust) to the hydrothermal solution would, on meeting Fe^{2+} at the margins of the growing mound, cause the reprecipitation (or migration) of an FeS front (51). Calculated using Geochemist's Workbench (35). (Redrawn from reference 16. Copyright 2006 Geological Society of America.)

Acetate, Amino Acids and Heterochiral Peptides

As we have seen, the likely first metabolic pathway driven by the disequilibrium between carbon oxides and hydrothermal hydrogen is toward the synthesis of acetate. In a notable experiment Huber and Wächtershäuser (*60*) generated high yields of activated acetate from CO and CH_3SH (methane thiol) in hydrothermal conditions previously synthesized from CO_2 and H_2S by Heinen and Lauwers (*61*). Yields were generally greatest in alkaline conditions. Nickel was the most effective catalyst, whether as (Fe,Ni)S, NiS or $NiSO_4$. In further experiments

Huber and Wächtershäuser (*62*) went on to reductively aminate alpha-carboxylic acids to a number of amino acids. The highest yields were restricted to pH conditions close to the pK_a of ammonium (9.25, Figure 7), i.e., between pH 8.5 and 9.5. In turn, peptides have been synthesized from amino acids using CO or COS as condensing agents. Again the best results were obtained between pH 8 and 10, close to the pK_a of the amino group of the amino acids in question (*63,64*). Similar high peptide yields were achieved at similar pH by Leman and coworkers (*65*) who used carbonyl sulfide as condensing agent (Figure 8).

In a finding of paramount importance to the debate on chirality and the origin of life, Huber and her coworkers (*64*) found that L-tyrosine immediately racemized on dimerization, indicating that appeals to a Universal bias in favor of L amino acids have no impact on subsequent handedness of peptides. The result bears on the theoretical modeling of the interaction of short peptides with both cations and anions. Milner-White and Russell (*34,66*) demonstrate theoretically that short achiral peptides sequester the anions significant to biosynthesis and electron transfer, viz., $HP_2O_7^{3-}$ and $[Fe_4S_4][CH_3S]_4^{2-}$, which are drawn to the δ^+ charges on the amines along the backbone of the peptide chain. In the relatively alkaline conditions present in the mound the chain would also sequester cations such as Ni^{2+} and Co^{2+} in forms that could encourage group transfer, condensations and reductions (*66*). These achiral peptides might also take over the role of cell wall and membrane (as amyloid) from the inefficient iron sulfide barriers (*66*). Thus a path from abiotic chemistry to an unregulated metabolism is now discernable. What is missing is a catalyst and code for synthesis and replication. This is an issue beyond the scope of this chapter though Copley and her coworkers (*41*) show the way. They model how particular alpha-keto acids might be aminated at the same dinucleotide site that happens to comprise the first two bases of the codon specifying that same amino acid (*41*). But returning to the supplies of energy, we now examine how the likely first metabolic pathways resulting from the various disequilibria obtained at a submarine alkaline spring are partly echoed in a modern comparable environment briefly described below.

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

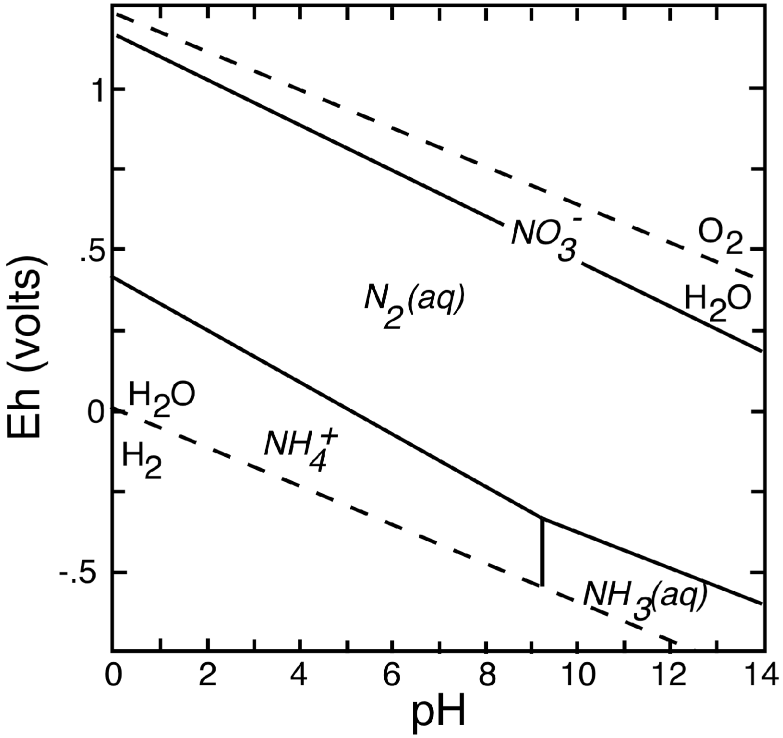


Figure 7. Pourbaix diagram to illustrate the stability fields of nitrogen species and in particular to show the NH_4^+/NH_3 equilibrium at pH 9.25 (ammonium's most interactive state) and 25°C. The NH_4^+/NH_3 equilibrium becomes more alkaline at lower temperature (19).

Microbiology of a Modern Submarine Alkaline Vent

Although the geochemistry of a modern alkaline vent and its environs such as Lost City (*67*) differs from Hadean examples, particularly with regard to the presence of dissolved oxygen and sulfate in the present-day ocean, it is instructive to consider the prokaryotes hosted there. A Methanosarcinales phylotype dominates at least the hotter portions of the Lost City chimneys (*68*). This methanogen occurs within biofilm tens of micrometres thick that probably also contains methanotrophs of similar genetic makeup (*69*). The outer, cooler portions of the chimneys harbor sulfide-oxidizing filamentous eubacteria (*68*), a class not to be expected in the Hadean. And the possible presence of sulfate-reducing bacteria (*69*), also absent in the Hadean, may explain the otherwise puzzling (to us) absence of homoacetogens. Presumably any would-be homoacetogens find it difficult to compete for hydrogen with sulfate-reducers in this environment (*70*). The high hydrogen content of the Lost City fluids (≤15 mM, *18*) fuels both the methanogens and any sulfate reducers, while formate may supply the carbon (*71*).

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

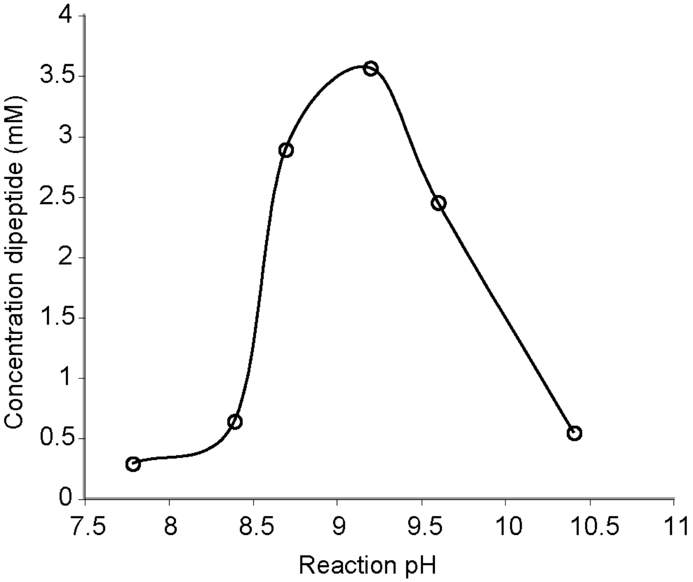


Figure 8. Dependence of the room-temperature formation of phenylalinine dipeptide on pH (cf. 62,63). (From reference 65. Copyright 2008 American Society for the Advancement of Science.)

Discussion

Autogenic theories on the origin of life have always been faced with an energy short-fall—how in the face of kinetic and thermodynamic inhibitions were the first pathways from simple one-carbon compounds to organic life negotiated? The protonmotive force that drove, and drives still, biosynthesis, first resulted from the pH gradient operating across a membrane. Thus it was an ambient feature of the geochemical conditions whereby the alkaline hydrothermal solution interfaced the mildly acidic early oceans across semipermeable and semiconducting inorganic barriers (*30*). Its operation has of course, since been refined and developed with the evolution of the rotor-stator ATPase but it does appear to have been required in some form just to get autogenic metabolism going (*7*). The thermal gradient is also significant. It had the potential to have concentrated the first products within pore spaces near the outer margins of the mound through thermal diffusion, and it could also have driven a convective polymerase chain reaction for the replication of nucleic acid polymers (*39*).

So metabolism was propelled into being on Earth ~4 billion years ago by chemical, electrochemical and thermal disequilibria. It emerged at the semi-lithified physical front separating an alkaline hydrothermal mound from the acidulous ocean (*72*). This is where chemical, redox, pH, and temperature gradients were at their steepest, yet were commensurate with the requirements of metabolic processes. The chemical gradients—CO_2 on the outside and H_2 on the inside—provided the ultimate chemical potential for their interaction, catalyzed within the metal sulfide compartment walls (*16,30*). The alkaline interior was conducive to acetate and amino acid synthesis and to the condensation of the former and polymerization of the latter (*60,62,63*).

Life per se might be said to have emerged once the biochemical reduction of the CO_2 dissolved in the Hadean Ocean through a coded and regulated metabolism had been achieved (*40*). Thereafter it developed in two channeled thermodynamically-driven processes, to acetogenesis and methanogenesis, genetically distinguished as processes mastered by the bacteria and the methanoarchaea respectively. Homoacetogenic bacteria and the methanoarchaea were the probable root organisms of the two prokaryote domains, viz., the bacteria and the archaea (*7*).

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

Acknowledgments

MJR's research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. He thanks Isik Kanik for help, encouragement and support.

References

1. Osborn, H. F. *The Origin and Evolution of Life: On the Theory of Action, Reaction and Interaction of Energy;* Charles Scribner's Sons: New York, 1917.
2. Leduc, S. *The Mechanism of Life;* Rebman Ltd: London, 1911.
3. Mereschkowsky, C. *Biol. Centralbl.* **1910,** *30,* 278–288 & 289–303 & 321–347 & 353–367.
4. Wicken, J. S. *Evolution, Information and Thermodynamics: Extending the Darwinian Program;* Oxford University Press: New York, 1985.
5. Crabtree, R. H. *Science* **1997,** *276,* 222.
6. Morowitz, H.; Smith, E. *Santa Fé Special Paper* **2006,** #06-08-029.
7. Martin, W.; Russell, M. J. *Phil Trans R. Soc. London (Ser.B)* **2007,** *362,* 1887–1925.
8. Shock, E. L. *Origins Life Evol. Bios.* **1992,** *22,* 67–107.
9. Li, Z.-X. A.; Lee, C.-T. A. *Earth Planet. Sci. Lett.* **2004,** *228,* 483–493.
10. Halevy, I.; Zuber, M. T.; Schrag, D. P. *Science* **2007,** *318,* 1903–1907.
11. Wood, B. J.; Bryndzia, L. T.; Johnson, K. E. *Science* **1990,** *248,* 337–345.
12. Delano, J. W. *Origins Life Evol. Bios.* **2001,** *31,* 311–341.
13. Touboul, M.; Kleine, T.; Bourdon, B.; Palme, H.; Wieler, R. *Nature* **2007,** *450,* 1206–1209.
14. Macleod, G.; McKeown, C.; Hall, A. J.; Russell, M. J. *Origins Life Evol. Biosph.* **1994,** *24,* 19–41.
15. Abrajano, T. A.; Sturchio, N. C.; Kennedy, B. M.; Lyon, G. L.; Muehlenbachs, K.; Bohlke, J. K. *Appl. Geochem.* **1990,** *5,* 625–630.
16. Russell, M. J.; Hall, A. J. *Geol. Soc. Am. Mem.* **2006,** *198,* 1–32.

17. Proskurowski, G.; Lilley, M. D.; Kelley, D. S.; Olson, E. J. *Chem. Geol.* **2006,** *229,* 331–343.
18. Früh-Green, G. L.; Kelley, D. D.; Bernascono, S. M.; *et al. Science* **2003,** *301,* 495–498.
19. Amend, J. P.; Shock, E. L. *FEMS Microbiol. Rev.* **2001,** *25,* 175–243.
20. Amend, J. P. In *Chemical Evolution II: From Origins of Life to Modern Society*. Am. Chem. Soc. 2008, this volume.
21. Shock, E. L. *Origins Life Evol. Bios.* **1990,** *20,* 331–367.
22. Fuchs, G. In *Autotrophic Bacteria;* Schlegel, H. G.; Bowien, B., Eds.; Science Tech.: Madison, WI, 1989**;** pp 365–382.
23. Thauer, R. K. *Microbiol.* **1998,** *144,* 2377–2406.
24. McCollom, T.; Seewald, J. S. *Geochim. Cosmochim. Acta* **2003,** *67,* 3645–3664.
25. Fuchs, G. In *Biology of the Prokaryotes*; Lengeler, J. W.; Drews, G.; Schlegel, H. G., Eds.; Blackwell Science: Oxford, United Kingdom, 1999; pp 163–186.
26. Grabarse, W.; Mahlert, F.; Duin, E. C.; Goubeaud, M.; Shima, S.; Thauer, R. K.; Lamzin, V.; Ermler, U. *J. Mol. Biol.* **2001,** *309,* 315–330.
27. Lolkema, J. S.; Chaban, Y.; Boekema, E. J. *J. Bioenerg. Biomembr.* **2003,** *35,* 323–335.
28. Kell, D. B. In *Bacterial Energy Transduction;* Anthony, C., Ed.; Academic Press: London, 1988; pp 429–490.
29. Volbeda, A.; Fontecilla-Camps, J. C. *Coord. Chem. Rev.* **2005,** *249,* 1609–1619.
30. Russell, M. J.; Hall, A. J. *J. Geol. Soc. Lond.* **1997,** *154,* 377–402.
31. Baltscheffsky, M.; Schultz, A.; Baltscheffsky, H. *FEBS Lett.* **1999,** *457,* 527–533.
32. Rouse, R. C.; Peacor, D. R.; Freed, R. L. *Am. Mineral.* **1988,** *73,* 168–171.
33. Popova, V. I.; Popov, V. A.; Sokolova, E. V.; Ferraris, G.; Chukanov, N. V. *N. Jb. Miner. Mh. Jg.* **2002.**
34. Milner-White, E. J.; Russell, M. J. *Origins Life Evol. Bios.* **2005,** *35,* 19–27.
35. Bethke, C. *Geochemical Reaction Modeling;* Oxford University Press: Oxford, United Kingdom, 1996.
36. Ragsdale, S. W. *Crit. Rev. Biochem. Molec. Biol.* **2004,** *39,* 165–195.
37. Seewald, J. S.; Zolotov, M. Y.; McCollom, T. *Geochim. Cosmochim. Acta* **2006,** *70,* 446–460.
38. Maden, B. E. H. *Biochem. J.* **2000,** *350,* 609–629.
39. Baaske, P.; Weinert, F. M.; Duhr, S.; Lemke, K. H.; Russell, M. J.; Braun, D. *Proc. Natl. Acad. Sci. USA* **2007,** *104,* 9346–9351.
40. Koonin, E. V.; Martin, W. *Trends Genetics* **2005,** *21,* 647–654.
41. Walker, J. C. G. *Origins Life Evol. Bios.* **1985,** *16,* 117–127.
42. Kasting, J. F.; Ackerman, T. P. *Science* **1986,** *234,* 1383–1385.
43. Yamagata, Y.; Wanatabe, H.; Saitoh, M.; Namba, T. *Nature* **1991,** *352,* 516–519.
44. Lowell, R. P.; Keller, S. M. *Geophys. Res. Lett.* **2003,** *30,* 1391.
45. Kump, L. R.; Seyfried, W. E. *Earth Planet. Sci. Lett.* **2005,** *235,* 654–662.
46. Mozgova, N.; Krasnov, S. G.; Battiyev, B. N.; Borodaev, Y. *Can. Mineral.* **1996,** *34,* 23–28.

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

47. Hem, S.; Makovicky, E.; Gervilla, F. *Can. Mineral.* **2001,** *39,* 831–853.
48. Zaikov, V. V.; Melekestseva, I. Yu. *Geologiya Rudnykh Mestorozhdenii* **2006,** *48,* 151-174.
49. Nekrasov, I. Y.-A.; Konyushok, A. A. *Mineralogicheskii Zhurnal* **1982,** *4,* 33–40.
50. Helz, G. R.; Miller, C. V.; Charnock, J. M.; Mosselmans, J. F. W.; Pattrick, R. A. D.; Garner, C. D.; Vaughan, D. J. *Geochimica Cosmochimica Acta* **1996,** *60,* 3631–3642.
51. Lennie, A. R.; Vaughan, D. J. In *Mineral Spectroscopy; a Tribute to Roger G. Burns;* Darby, D.M.; McCammon, C.; Schaefer M.W., Eds; Geochem. Soc. Spec. Publ. 5; 1996; pp 117–131.
52. Rickard, D.; Luther, G. W. *Chem. Rev.* **2007,** *107,* 514–562.
53. Rickard, D.; Butler, I. B.; Olroyd, A. *Earth Planet. Sci. Lett.* **2001,** *189,* 85–91.
54. Filtness, M. J.; Butler, I. B.; Rickard, D. *Trans. Instn Min. Metall.* **2003,** *112B,* 171–172.
55. McCollom, T.; Seewald, J. S. *Geochim. Cosmochim. Acta* **2003,** *67,* 3625–3644.
56. Lang, S. Q.; Butterfield, D. A.; Lilley, M. D.; Johnson, H. P.; Hedges, J. I. *Geochim. Cosmochim. Acta* **2006,** *70,* 3830–3842.
57. Blöchl, E.; Keller, M.; Wächtershäuser, G.; Stetter, K. O. *Proc. Natl. Acad. Sci. USA* **1992,** *89,* 8117–120.
58. Brandes, J. A.; Boctor, N. Z.; Cody, G. D.; Cooper, B. A.; Hazen, R. M.; Yoder, H. S. *Nature* **1998,** *395,* 365–367.
59. Dörr, M.; Käßbohrer, J.; Grunert, R.; Kreisel, G.; Brand, W. A.; Werner, R. A.; Geilmann, H.; Apfel, C.; Robl, C.; Weigand, W. *Angew. Chem. Int. Ed.* **2003,** *42,* 1540–1543.
60. Huber, C.; Wächtershäuser, G. *Science* **1997,** *276,* 245–247.
61. Heinen, W.; Lauwers, A. M. *Origins Life Evol. Bios.* **1996,** *26,* 131–150.
62. Huber, C.; Wächtershäuser, G. *Tetr. Lett.* **2003,** *44,* 1695–1697.
63. Huber, C.; Wächtershäuser, G. *Science* **1998,** *281,* 670–672.
64. Huber, C.; Eisenreich, W.; Hecht, S.; Wächtershäuser, G. *Science* **2003,** *301,* 938–940.
65. Leman, L.; Orgel, L.; Ghadiri, M. R. *Science* **2004,** *306,* 283–286.
66. Milner-White, E. J.; Russell, M. J. *Biol. Direct.* (in review).
67. Kelley, D. S.; Karson, J. A.; Blackman, D. K. et al. *Nature* **2001,** *412,* 145–149.
68. Schrenk, M. O.; Kelley, D. S.; Bolton, S. A.; Baross, J. A. *Env. Microbiol.* **2004,** *6,* 1086–1095.
69. Brazelton, W. J.; Schrenk, M. O.; Kelley, D. S.; Baross, J. A. *Appl. Envir. Microbiol.* **2006,** *72,* 6257–6270.
70. Krumholz, L. R.; Harris, S. H.; Tay, S. T.; Suflita, J. M. *Appl. Envir. Microbiol.* **1999,** *65,* 2300-2306.
71. Haggerty, J. A.; Fisher, J. B. *Leg125. Proc, Ocean Drilling Prog., Scientific Results, 125:* College Station, TX, **1992**; pp 387–395.
72. Russell, M. J.; Daniel, R. M.; Hall, A. J.; Sherringham, J. A. *J. Mol. Evol.* **1994,** *39,* 231–243.

Downloaded by STANFORD UNIV GREEN LIBR on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch003

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Chapter 4

Energetics of Biomolecule Synthesis on Early Earth

Jan P. Amend[1] and Tom M. McCollom[2]

[1]Department of Earth and Planetary Sciences, Campus Box 1169, Washington University, St. Louis, MO 63130, USA
[2]Laboratory for Atmospheric and Space Physics, Campus Box 392, University of Colorado, Boulder, CO 80309-0392, USA

Among the most plausible environments for the origin of life are marine hydrothermal systems, where geochemical energy sources and refugia from sterilizing meteorite impacts would have been plentiful. Here, values of Gibbs energy were calculated for the formation of individual cellular building blocks from inorganic reactants. In our model, corresponding redox reactions occurred at the interface between two end-member fluids─low temperature (25 °C), mildly acidic (pH 6.5), relatively oxidized (E_h -0.30 mV) seawater and moderately hot (140 °C), alkaline (pH 9), reduced (E_h -0.71 mV) hydrothermal vent fluid. The thermodynamic calculations demonstrate that biomass synthesis is most favorable at moderate temperatures, where the energy contributions from HCO_3^- and H^+ in seawater coupled to the reducing power in hydrothermal fluid are optimized. The models further show that the net synthesis of cellular building blocks may yield small amounts of energy over the ranges of temperature and chemical composition investigated. This is counter to conventional wisdom for anabolic processes and lends further support to marine hydrothermal systems as particularly favorable sites for the emergence of life.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Background

There is accumulating evidence that life emerged on Earth during the late Hadean era (~4.2-3.8 Ga), and that microorganisms thrived by the early Archean era (3.8-3.4 Ga). In one study, isotopically light carbonaceous inclusions in phosphate mineral (apatite) grains from ~3.8 Ga sediments in West Greenland were interpreted as robust evidence of biological activity (*1*). Carbon isotopes were also used as a biosignature of planktonic organisms in slightly younger (~3.7 Ga) pelagic sediments from Greenland (*2*). These data further suggested that oxygenic photosynthesis had evolved by this time, a conclusion which is vigorously debated. In still younger rocks, the ~3.5 Ga Apex chert in the Warrawoona Group of Western Australia, putative microfossils were interpreted as oxygen-producing cyanobacteria (*3, 4*). However, in a reinterpretation of these cherts, the microfossil-like structures were labelled as artifacts from amorphous graphite (*5*), and it was shown that Raman spectroscopy, the analytical tool used by Schopf and co-workers, cannot unambiguously identify biogenicity (*6*). Although organic biomarkers, ancient soils, and stromatolites provide strong evidence for the advent of oxygenic photosynthesis by ~2.5-2.7 Ga, these interpretations also are not entirely without controversy (*7-10*).

Regardless of when oxygenic photosynthesis evolved, it is widely accepted that it was preceded by its anoxygenic counterpart. Evidence for this comes from analyses of numerous photosynthesis genes and whole genomes of all five recognized groups of photosynthetic bacteria—cyanobacteria, purple bacteria, green sulfur bacteria, green non-sulfur bacteria, and heliobacteria (*11-13*). In addition, two geochemical studies suggest that H_2-based carbon fixation via anoxygenic photosynthesis occurred in microbial mats that are preserved in the ~3.4 Ga Buck Reef Chert, South Africa (*14, 15*). There, the evidence includes the presence and absence of certain trace minerals (e.g., iron-carbonates and iron-oxides), and specific rare earth element patterns, which imply an anoxic water column despite the apparent presence of phototrophs.

Predating both oxygenic and anoxygenic phototrophy appears to be chemotrophy, and in particular, chemolithoautotrophy (in this metabolic process, organisms utilize inorganic energy sources, in part, to synthesize their own biomass from CO_2 and other inorganic carbon sources). As noted by Martin and Russell (*16*), Mereschkowsky (*17*) was the first to argue that the earliest organisms were anaerobic and autotrophic, with the ability to synthesize carbohydrates without the aid of chlorophyll. A chemolithoautrophic origin of life is now supported by biochemical, geochemical, and phylogenetic evidence. For example, Martin and Russell (*16*) and Russell and Hall (*18*) built on a study by Fuchs (*19*) to argue for an ancient CO_2-fixation pathway, perhaps similar to the acetyl-CoA (Wood-Ljungdahl) pathway. Support comes from this pathway's occurrence in anaerobic and thermophilic bacteria and archaea, its low energy requirements, and the positions of acetogens and methanogens deep in the tree of life. Geochemical considerations favor chemolithoautotrophy, especially in hydrothermal systems, on energetic grounds. Numerous studies calculated the thermodynamic drive for abiotic organic synthesis and the energy yields from chemolithotrophy (*20-26*). Phylogenetic arguments include those of Pace (*27, 28*), who constructed a global tree of life from which he concluded

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

that the last common ancestor, and by inference the earliest organism, was chemotrophic, autotrophic, and thermophilic. This is the most parsimonious interpretation of the organisms that occupy the deepest and shortest branches in the archaea and bacteria domains. The majority of subsequent investigations have confirmed or complemented these views, concluding consequently that organoheterotrophy—the reliance on organic compounds as energy and carbon sources—evolved after lithoautotrophy.

Ever since their discovery, marine hydrothermal systems have been put forth as possible environments for the origin of (thermophilic) life (*29-33*). In addition to the phylogenetic and thermodynamic inferences mentioned above, support for this theory also comes from planetary arguments. Until ~3.8 Ga, the Earth was subject to heavy bombardment by massive meteorites that could have sterilized all near-surface environments. It has been proposed that deep-sea hydrothermal systems could have served as refugia for the survival and evolution of early life (*34-36*). It perhaps should be noted, however, that some researchers continue to argue against hydrothermal system theories and prefer the primordial ocean as the spring of life, with heterotrophy as the first metabolic strategy (*37*). Nevertheless, based on the abundance of evidence from numerous and disparate sources, we adopt as most likely that the first organism was a chemolithoautotroph, and that its niche was a marine hydrothermal system.

The focus then shifts to describing the physicochemical properties of that system. Because metabolic processes are primarily electron transfer processes, redox disequilibria must have existed in such a system to facilitate the transition from a sterile, prebiotic world to one inhabited by single-celled life forms. A reaction zone can be envisioned where hot, chemically reduced, perhaps slightly alkaline hydrothermal fluids mixed with cooler, more oxidized, arguably acidic seawater to generate a redox front (*18*). It is at such an interface that inorganic compounds may have reacted to form simple organic molecules, which then polymerized and ultimately led to something that resembled a compartmentalized cell and the origin of life. In an effort to better understand potential metabolic reactions in such an environment, we calculate the Gibbs energy of reaction (ΔG_r) for the synthesis of individual molecular building blocks (amino acids, nucleotides, fatty acids, saccharides, amines) that constitute prokaryotic biomass. The reactants are inorganic compounds, and the energetics apply to a hydrothermal solution with seawater and vent fluid end-members.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Mixing of Seawater and Vent Fluid

Rapid mixing of two chemically distinct fluids yields a solution that is, at least temporarily, out of thermodynamic equilibrium. Here, we propose the formation of a hydrothermal solution that is the product of late Hadean vent fluid mixing into late Hadean seawater. In this mixed solution, disequilibria among redox sensitive compounds represent an energetic drive for the abiotic synthesis of organic compounds. Similar mixing scenarios have been used effectively to model the energetic yields of chemolithotrophic metabolisms on Earth (*21*), Mars (*38, 39*), and Jupiter's moon Europa (*40*), and to evaluate the energetics of organic synthesis (*23, 41*), including amino acids (*22, 42*). The models first require that the compositions of the two end-member fluids be defined.

Composition of Late Hadean Seawater

Defining a representative composition for the ocean end-member in the models is particularly challenging because nearly every aspect of the physical and chemical state of the coupled ocean/atmosphere system on the early Earth (temperature, pH, elemental composition, salinity, oxidation state, etc.) is poorly understood at present and continues to generate vigorous debate. We adopt an ocean composition that appears consistent with the currently available constraints. The composition of the model late Hadean seawater used in this communication is given in Table I.

There appears to be a general consensus that the atmosphere prior to the advent of life was mildly reducing, dominated by N_2 and CO_2, with minor amounts of reduced gases such as H_2 and CH_4, and very little or no O_2 (*43, 44*). However, the actual abundances of these components in the early atmosphere, and consequently in the early ocean, remain uncertain. Russell and colleagues have advocated a moderately acidic early ocean, with the acidity attributable to the formation of carbonic acid in equilibrium with an atmosphere containing 1-10 bar CO_2 (*18*). Elevated CO_2 has also been suggested as a means of providing greenhouse warming to counter the faint young sun (*43, 45*). However, geologic observations and early atmosphere models suggest that several bars of CO_2 might not have been tenable given geologic constraints (*7*). Methane (CH_4) also has been suggested as an alternative to CO_2 as a key greenhouse gas for the early Earth (*46, 47*), but perhaps a CO_2-enriched atmosphere gave way to a CH_4-enriched atmosphere only after the onset of biological methanogenesis (*44*). For this study, we adopt a late Hadean ocean in equilibrium with an atmosphere that contains elevated, but not extreme, levels of CO_2 (0.2 bar). This leads to a mildly acidic ocean with pH equal to 6.5.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table I. Compositions of End-Member Fluids Used in the Mixing Calculations

	Seawater	*Hydrothermal Fluid*
T (°C)	25	140
pH	6.5	9.0
E_h (mV)	-0.30	-0.71
$H_{2(aq)}$	0.00054	16.
ΣCO_2	22.7	0.01
$O_{2(aq)}$	10^{-9}	0.
Na	464.	607.
Cl	540.	650.
Ca	14.2	20.
Mg	31.1	0.001
Fe	0.12	0.
K	0.1	10.
$CH_{4(aq)}$	0.	2.0
ΣH_2S	0.0002	0.1
SO_4^{2-}	0.1	0.
$SiO_{2(aq)}$	0.11	0.005

NOTE: All concentrations in *m*mol kg^{-1}.

It has been argued, largely on the basis of oxygen isotopes, that the Archean ocean had temperatures substantially elevated relative to the modern ocean (i.e., 45-70ºC) (*48, 49*). Conversely, others have argued that such elevated temperatures may not be consistent with the geologic evidence, and that the isotopic observations can be accounted for by other explanations (*47*). In the absence of more definitive constraints, we adopt a conservative temperature of 25 ºC. Further, late Hadean seawater is taken here to have had concentrations of dissolved Na and Cl similar to modern seawater, and concentrations of Ca, Mg, and Fe set by saturation with respect to the common minerals calcite ($CaCO_3$), magnesite ($MgCO_3$), and siderite ($FeCO_3$), respectively. Similarly, aqueous SiO_2 and H_2S are set by saturation with respect to quartz (SiO_2) and pyrite (FeS_2), respectively. Tian et al. (*50*) recently suggested that the H_2 level in the early atmosphere was 0.1 bar or higher, but Catling and Claire (*44*) argue that an atmospheric mixing ratio on the order of 10^{-3} was more likely. Accordingly, our model sets the early ocean in equilibrium with 0.001 bar of atmospheric H_2, leading to an aqueous concentration of 0.54 μmol kg^{-1}. We further assume that equilibrium with dissolved H_2 controls the oxidation state of dissolved Fe. Sulfur isotope signatures from the early Earth constrain O_2 levels to below 10^{-5} bar (*51*), but because the precise level is unknown, we assumed a dissolved concentration of 10^{-12} mol kg^{-1}. Sulfate levels in the early ocean were apparently <0.2 *m*mol kg^{-1} (*52*), and we assign it a nominal concentration of 0.1 *m*mol kg^{-1}.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Composition of Late Hadean Vent Fluid

It has been proposed that the origin of life occurred where moderate temperature (<~150 °C) hydrothermal fluids discharged from serpentinized ultramafic rocks (*16, 18, 53*). Owing to chemical reactions taking place during the reaction of ultramafic rocks (<45% SiO_2 and high MgO and FeO content) with water during serpentinization, fluids discharged from serpentinites at moderate to low temperatures are alkaline with high levels of reduced compounds (e.g., H_2, CH_4). The production of H_2 in this process can be portrayed by the general reaction

$$\underset{\text{(Olivine)}}{(Mg,Fe)_2SiO_4} + aH_2O \rightarrow \underset{\text{(Serpentine)}}{x(Mg,Fe)_3Si_2O_5(OH)_4} + \underset{\text{(Brucite)}}{yMg(OH)_2} + \underset{\text{(Magnetite)}}{zFe_3O_4} + zH_2. \qquad (1)$$

where the stoichiometric coefficients (*a, x, y,* and *z*) and mineral compositions depend on several factors including temperature, pressure, and water-rock ratio. On the modern Earth, serpentinites form primarily where ultramafic rocks from the mantle are transported to near surface environments by tectonic processes, but on the early Earth, ultramafic rocks would have been more widespread owing to the occurrence of komatiitic lava flows (*54, 55*).

The Lost City system in the Mid-Atlantic (*56, 57*) represents a modern analog of the alkaline hydrothermal vents proposed by Russell and colleagues as the site of the origin and early evolution of life. At Lost City, warm (<30-90 ºC), alkaline (pH = 9-11) hydrothermal fluids discharge at the seafloor, where they form chimneys composed of carbonate minerals and brucite ($Mg(OH)_2$). The fluids contain high levels of H_2 generated by the serpentinization process (reaction 1) as well as elevated levels of CH_4. Dissolved inorganic carbon in the fluids is very low, owing to precipitation of carbonate minerals in the subsurface at the strongly alkaline pH.

The composition of the end-member late Hadean model vent fluid used in our mixing calculation is based on the measured composition of the Lost City vent fluids (Table I). Although the highest seafloor fluid temperatures measured there are only ~90 ºC, the fluid chemistry suggests that the composition is controlled by water-rock reactions at higher temperatures (*58, 59*). Based on H and O isotopes, Proskurowski et al. (*59*) estimate the temperature in the subseafloor reaction zone to be <150 ºC; therefore, we adopt a temperature of 140 ºC for our end-member hydrothermal vent fluid. The pH is set to 9, which is at the lower end of the range measured at Lost City; the more alkaline pH values recorded in the vent fluids require respeciation of the fluid at temperatures below 90 ºC. Concentrations of H_2, CH_4, and Ca in the fluid are based on reported concentrations at Lost City (*57*). Concentrations of HCO_3^- and dissolved Mg are set by saturation with calcite and brucite, respectively. Concentrations of other components (K, Fe, SiO_2, SO_4^{2-}, H_2S, and NH_4^+) in the fluids at Lost City have not been reported, and so are assigned nominal concentrations here.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Mixing Calculations

The composition of the fluid during mixing of late Hadean seawater with late Hadean vent fluid was calculated using the computer program EQ3/6 (*60*). All calculations were performed at 250 bar to simulate the elevated pressures of the deep sea. The thermodynamic database required for EQ3/6 was generated using the SUPCRT92 computer program (*61*), and includes thermodynamic data for minerals from Helgeson et al. (*62*) and for aqueous species and complexes from Shock and Helgeson (*63*), Shock et al. (*64, 65*), Sverjensky et al. (*66*), and McCollom and Shock (*21*).

The model starts with 1 kg of vent fluid at 140 ºC, and then adds small amounts of 25 ºC seawater in increments, continuously recalculating the composition and chemical speciation of the mixed fluid after each step. The temperature of the mixed hydrothermal solution is assumed to scale linearly with the proportions of the end-member fluids. It was presumed that all redox reactions were kinetically inhibited on the timescales of mixing. Minerals were allowed to precipitate during mixing, except that precipitation of graphite and dolomite were presumed to be kinetically inhibited. Mixing resulted in the precipitation of 141 g of carbonate [$(Ca_{0.46}Mg_{0.47}Fe_{0.07})CO_3$], together with trace amounts (<<1 g) of brucite ($Mg(OH)_2$) and monticellite ($CaMgSiO_4$) from each kg of vent fluid mixed with seawater, which primarily occurs at high seawater:hydrothermal fluid mixing ratios.

The changing composition of the hydrothermal solution during mixing is shown in Figure 1, where it can be seen that the pH drops substantially from 9 to ~6 during initial mixing. Since redox reactions were presumed to be kinetically inhibited, H_2 concentrations show a gradual decrease with increasing seawater:hydrothermal fluid ratio (SW:HF), reflecting conservative mixing. Owing to the combined effects of decreasing H_2 and pH, the electrode (or reduction) potential (E_h) of the fluid increases sharply from -0.71 mV at 140 °C to >-0.5 mV at 125 °C during initial mixing. The E_h then increases more gradually with increasing SW:HF as temperatures drop from 125 to 25 °C. Other components of the fluid exhibit gradual changes in composition with mixing, reflecting the contributions from the two end-members and small

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

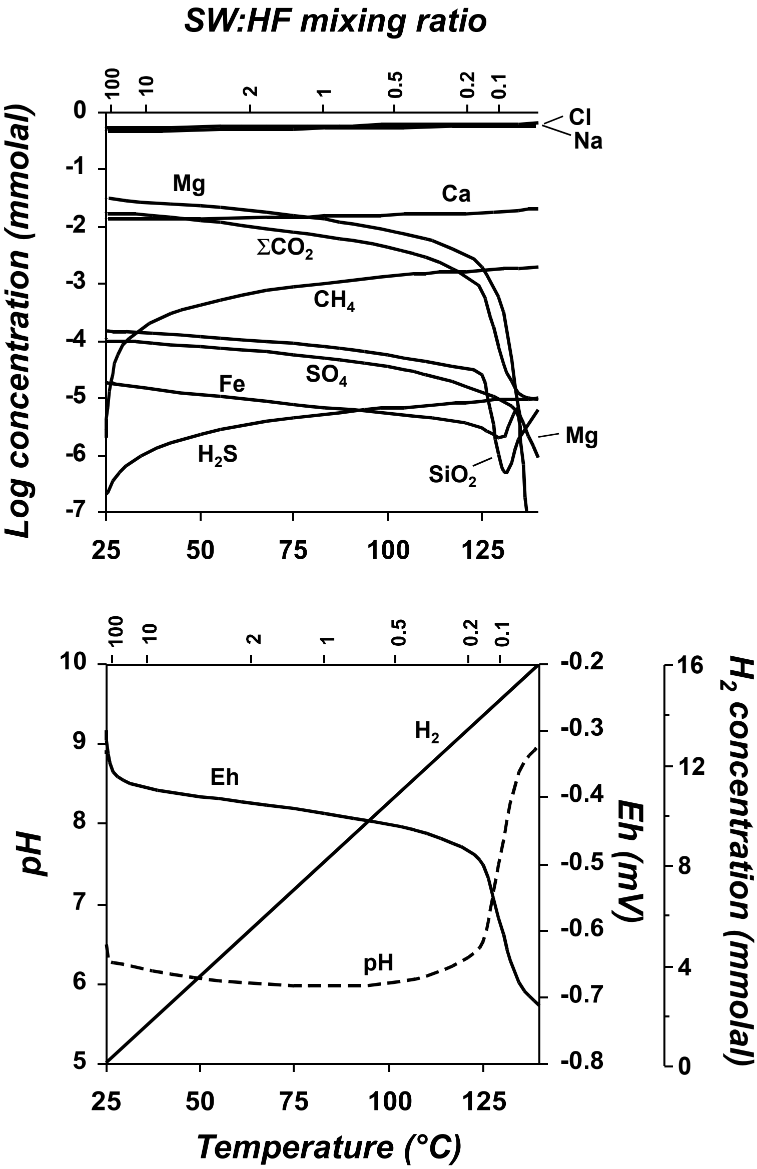


Figure 1. Calculated composition of the fluid during mixing of alkaline hydrothermal vent fluid with seawater on the early Earth. Compositions are shown as a function of temperature and the equivalent seawater:hydrothermal fluid (SW:HF) mixing ratio.

losses, particularly of Ca, Mg, and dissolved CO_2, from the fluid owing to mineral precipitation. Because there are few constraints on the amount of ammonia in seawater or hydrothermal fluid on the early Earth, the calculations assume a constant, conservative concentration (10^{-8} *m*mol kg^{-1}) in the mixed solution at all temperatures.

Energy Calculations

The energetics are calculated for the synthesis of biomolecules that make up a representative prokaryotic cell, starting with inorganic constituents. This approach builds on pioneering work by Morowitz (*67*) to assess the energy costs of anabolism in autotrophs, later studies by Stouthamer and Bettenhaussen (*68*) and Battley (*69-71*) to evaluate these costs in heterotrophs, and efforts by Amend and Shock (*22*) to model hydrothermal amino acid synthesis. Specifically, our computational method mimics that in our earlier work on biomass synthesis in present-day low temperature oxic and anoxic environments (*72*). In that study, prokaryotic biomass was divided into the individual molecular building blocks (amino acids, nucleotides, fatty acids, saccharides, amines) that comprise proteins, nucleic acids, membranes, cell walls, and other major cellular components. Values of ΔG_r are then calculated for the synthesis of these building blocks from HCO_3^-, NH_4^+, H_2S, HPO_4^{2-}, H^+, and electrons (e^-) at the temperature, pressure, and chemical composition that may have existed in a mixing zone of late Hadean seawater and hydrothermal vent fluid. Consistent with the laws of thermodynamics, the energetics are path-independent, which translates to viewing the cell as a 'black box' where inorganic precursors are converted to biomolecules without any attempt (or need) to describe the particular anabolic pathways. This can be represented by a chemical reaction with the generic form

$$aHCO_3^- + bNH_4^+ + cH_2S + dHPO_4^{2-} + eH^+ + fe^- \rightarrow u\text{Alanine} + v\text{AMP}^{2-} + w\text{Palmitate} + x\text{Glucose} + y\text{Putrescine} + \ldots + zH_2O, \quad (2)$$

where the reaction products represent the cellular building blocks.

The biomass synthesis reactions are written here in the biological convention as half-cell reactions with the hypothetical free electron (e^-). As discussed in detail in McCollom and Amend (*72*), this is thermodynamically equivalent to writing reactions with dissolved H_2, as long as the reaction

$$2H^+ + 2e^- \rightarrow H_2(aq) \quad (3)$$

is presumed to be in equilibrium (which is generally the case). In the half-cell reactions, the electron represents the 'reducing power' to convert the oxidized carbon in HCO_3^- to the reduced carbon in biomolecules. The activity of e^- (a_{e-}) can be directly related to E_h through the relation

$$\ln a_{e-} = -F\, E_h/RT, \quad (4)$$

where F stands for the Faraday constant, and R and T denote the gas constant and the temperature in Kelvin, respectively.

As representative prokaryote biomass, we use the composition of *Escherichia coli* (Table II) given by Battley (*69*). The chemical make-up of *E. coli* is much better known than that of any arguably more appropriate, deeply-branching archaeon or bacterium, but differences in biomass composition among microorganisms are probably of secondary importance relative to the bulk

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

chemistry of the system. And, as noted in McCollom and Amend (*72*), these calculations can be easily tailored to other system and/or biomass compositions as more information becomes available.

The energetics of the synthesis reactions were computed with the equation

$$\Delta G_r = \Delta G_r^o + RT \ln Q_r \,, \tag{5}$$

where ΔG_r, R, and T are as stated above, ΔG_r^o denotes the standard Gibbs energy of reaction r, and Q_r stands for the activity product, which can be calculated with the equation

$$Q_r = \Pi a_i^{\nu_{i,r}} \,, \tag{6}$$

where a_i represents the activity of species i and $\nu_{i,r}$ the stoichiometric reaction coefficient of that species in reaction r, which is negative for reactants and positive for products. In other words, the Q-term takes into account the chemical composition of the environment (here, the mixture of late Hadean seawater and hydrothermal vent fluid) and of the biomass (here, the intracellular concentrations of the biomolecular building blocks). The ΔG_r^o-term, on the other hand, takes into account the standard state Gibbs energies of each species in a reaction at the temperature and pressure of interest, which can be computed with the equation

$$\Delta G_r^o = \sum \nu_{i,r} \Delta G_i^o \,. \tag{7}$$

It follows from equations (5)-(7) that in order to estimate the amount of energy required to synthesize the cellular building blocks from inorganic precursors, we need: a) values of ΔG_i^o at the appropriate temperatures and pressures for all compounds involved in the synthesis reactions, b) estimates of the mixed hydrothermal fluid composition (discussed above), and c) representative biomass composition (here, that of *E. coli* given in Table II). The necessary values of ΔG_i^o were calculated with the computer code SUPCRT92 (*61*) using standard state thermodynamic properties and revised Helgeson-Kirkham-Flowers equation of state parameters published in the literature, or, in several cases, generated in this study. Values of ΔG_i^o (Table III) were computed at 250 bar and temperatures to 150 °C. These values of ΔG_i^o were then used with equation (7) to compute values of ΔG_r^o for the synthesis reactions at 250 bar and temperatures up to 150 °C, which are given in Table II.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Results and Discussion

Using the fluid compositions determined with the mixing models, the energetics (ΔG_r) of the biomass synthesis reactions shown in Table II were calculated with equation (5) for late Hadean mixed hydrothermal solutions for temperatures up to 125 °C. Results of these calculations are shown in Tables IV and V for temperatures of 125, 100, 75, and 50 ºC, corresponding to seawater-hydrothermal fluid mixing ratios of 0.15, 0.54, 1.3, and 3.6, respectively. Also shown are calculations at 25 ºC, corresponding to biomass synthesis at the assumed conditions of late Hadean seawater. The values of ΔG_r for the synthesis of 20 amino acids, 8 nucleotides, 5 fatty acids, 8 saccharides, and 4 amines, which are the monomeric constituents of model prokaryotic biomass, are given in Table IV in units of kilojoule per mole of monomer. In Table V, these values are converted to units of joule per gram of biomass (dry weight), which is here defined as the sum of the monomers. Values of ΔG_r (in J/g cells) for total amino acids, total nucleotides, total fatty acids, total saccharides, total amines, and total cell biomass are also listed in Table V and, in addition, are plotted in Figure 2.

It can be seen in Tables IV and V and in Figure 2 that the energetics for the monomer synthesis reactions differ substantially at different temperatures. This is due, on the one hand, to the temperature dependence of the ΔG_r^o -term, and on the other hand, to the compositional variation expressed in the Q_r-term. The composition of the mixed hydrothermal solution varies with temperature, because the end-member fluids (seawater and hydrothermal fluid) differ in temperature and chemical composition (Figure 1). Note that values of ΔG_r at 25

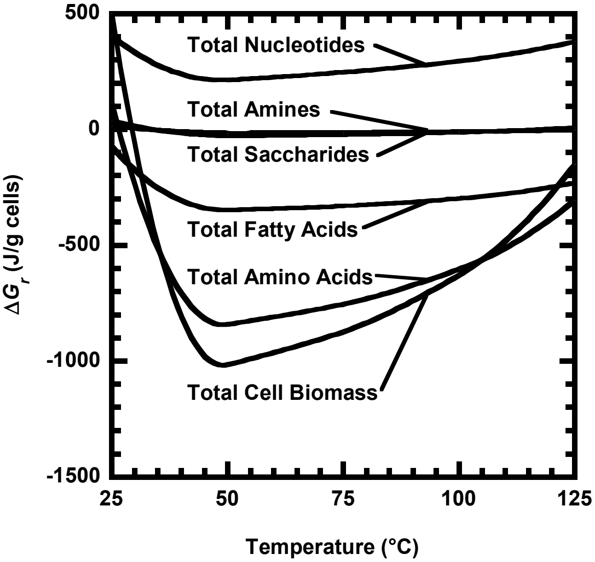


Figure 2. Calculated Gibbs energies of reaction as a function of temperature for molecular building blocks of cells and for total biomass.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

°C represent those in pure Hadean seawater, while those at 125 °C, for example, represent a SW:HF ratio of 0.15. ΔG_r^o and Q_r are both temperature dependent, but at low temperatures, differences in chemical composition (the Q_r-term) play a far greater role in the energetics (ΔG_r) than differences in the standard state properties (the ΔG_r^o-term). This can be seen by comparing values of ΔG_r^o in Table II with values of ΔG_r in Table IV. As an example, ΔG_r^o of alanine synthesis at 25 and 50 °C differ by 7 kJ/mol, but the corresponding difference in ΔG_r is 114 kJ/mol; the Q_r-term accounts for the difference between these two, or 107 kJ/mol. It should be pointed out, however, that the picture is quite different when considering larger temperature differences. If we again use the example of alanine synthesis, but this time at 25 and 125 °C, we note that the difference in ΔG_r^o is 35 kJ/mol, and the difference in ΔG_r is 55 kJ/mol; the Q_r-term accounts for only 20 kJ/mol. The same general trends hold for the other synthesis reactions as well.

Close inspection of Tables IV and V and Figure 2 reveals that the reaction energetics for all biomass monomers are most favorable in the mixed hydrothermal solution at moderate temperatures—here, this is best represented by 50 °C. At 25 °C (i.e., pure Hadean seawater), there is not enough reducing power to synthesize organic compounds from HCO_3^-, and values of ΔG_r^o are less negative than at higher temperatures. At 125 °C, electron donors from the hydrothermal fluid are plentiful, and values of ΔG_r^o are more negative, but these effects on ΔG_r are partially offset by the corresponding effects of low HCO_3^- concentrations and higher pHs. At 50 °C, the energy contributions from HCO_3^- and H^+ in seawater and from reducing power in the form of electrons in the hydrothermal fluid are optimized, resulting in the most negative Gibbs energy values. That said, the specific temperature (here, 50 °C) should not be overinterpreted. It merely indicates that a moderately elevated temperature relative to that of seawater, but with sufficient hydrothermal contribution for reducing power, is thermodynamically most favorable for the biomass synthesis scenario outlined here.

Further inspection of these results, best seen in Figure 2, shows that over the entire temperature range investigated, the synthesis of nucleotides is endergonic ($\Delta G_r > 0$) and only moderately temperature dependent; the synthesis of amines and saccharides occurs at or near equilibrium ($\Delta G_r = 0$); and that of fatty acids is exergonic ($\Delta G_r < 0$) and also only moderately temperature dependent. Amino acid synthesis, however, is endergonic only at very low temperatures; it is highly exergonic over most of the temperature range, and strongly temperature dependent. Note also in Figure 2 that the energetics of synthesizing total cell biomass (the sum of all monomers) mimics that of total amino acids. ΔG_r is positive only at low temperatures, minimizing at 50 °C, where the synthesis reactions are highly exergonic. With increasing temperature above 50 °C, ΔG_r increases, but remains negative even at the highest temperatures investigated.

The model results have two particularly significant implications for the origin and early evolution of life on Earth. First, the calculations show that in the transition from the prebiotic world to the biotic world, the energetics of

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

synthesizing the monomeric constituents of biomass are not only most favorable in a moderate temperature zone where seawater and hydrothermal fluid mix, but that, in fact, small amounts of chemical energy may be liberated by these processes. Second, the model results imply that once life got started in such environments, the proliferation of cells would have required relatively little energy input. Consequently, moderately thermophilic organisms that carried out methanogenesis or sulfur reduction, as just two of several metabolic examples, could have thrived in such hydrothermal systems on very anemic net energy sources.

The results of our computations are, of course, dependent on the assumed compositions of Hadean hydrothermal fluid and seawater, and will vary somewhat if the compositions of these fluids on the early Earth differ from those used here. However, because hydrothermal fluid compositions in serpentinizing systems are predominantly controlled by reactions with minerals in the subsurface and there is no reason to believe that the composition of ultramafic rocks on the early Earth were substantially different from those today (*73*), the composition of fluids discharged from serpentinites has probably changed very little over the past 4 billion years. One aspect of the assumed hydrothermal fluid composition, however, that may warrant further exploration is the level of dissolved inorganic carbon (DIC).

At Lost City, DIC levels (primarily as CO_3^{2-}) are very low. Seawater HCO_3^- is apparently quantitatively removed in the Lost City system by precipitation of carbonate minerals during circulation through the crust, and the dissolved carbon in the fluid, predominantly in the form of CH_4, is derived from dissolution and reduction of mantle CO_2 (*59, 74*). It is conceivable that DIC levels in serpentinite-hosted hydrothermal fluids on the early Earth may have been higher than in the Lost City system. Indeed, DIC levels in some terrestrial serpentinite-hosted alkaline springs are substantially higher than at Lost City (*75*), although the abundance of dissolved carbon in these groundwater-fed systems may have been influenced by interaction with carbonate-rich rocks in the subsurface.

If present at anything above trace levels, DIC in moderate- to low-temperature serpentinite-hosted hydrothermal fluids would be in disequilibrium with the amount of H_2 present in such highly reducing systems. In this case, Gibbs energies of biomolecule-forming reactions would be exergonic ($\Delta G_r < 0$) even in the endmember hydrothermal fluid, and mixing with seawater would not necessarily be required to provide a thermodynamic impetus for synthesis of biomolecules. Higher hydrothermal DIC contents would therefore make the thermodynamics of biomass synthesis more favorable at the high temperarture end of the calculations shown here.

Conclusions

It is a basic tenet of biochemistry that anabolism—the synthesis of cellular components from simple precursors—requires energy, and further, that this requisite energy derives from catabolism—the biochemical breakdown of organic or inorganic compounds in electron transfer processes. That energy

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

must be invested in biomass synthesis is largely based on observations made on organisms inhabiting the oxic and suboxic ecosystems that are pervasive on or near the surface of the modern Earth. The modelling efforts described here demonstrate that in strongly reducing environments where H_2 is abundant, such as occurs at sites of hydrothermal venting from ultramafic rocks, the energy requirements for biomass synthesis should be substantially lower. In some cases, the conversion (reduction) of bicarbonate to biomolecules may yield small amounts of additional metabolic energy to the organism. Based in large part on these energetic advantages in biomass synthesis, hydrothermal systems are viewed as particularly favorable sites for the origin of life. Perhaps, life got its start in the late Hadean at a temperature, pH, and redox front where biomass synthesis was energetically optimized. The additional challenges of investing metabolic energy in biomass synthesis may have occurred only as subsequent evolution enabled organisms to occupy an evermore geochemically diverse array of environments.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table IIa. Biomass Composition and Synthesis Reactions

Compound	*Amt.*[a] *[mg (g cells)$^{-1}$]*	*Biomass Synthesis Reaction*
Alanine	48.4	$3HCO_3^- + NH_4^+ + 14H^+ + 12e^- = C_3H_7NO_2 + 7H_2O$
Arginine$^+$	49.0	$6HCO_3^- + 4\ NH_4^+ + 25\ H^+ + 22e^- = C_6H_{15}N_4O_2^+ + 16H_2O$
Asparagine	30.3	$4HCO_3^- + 2\ NH_4^+ + 14H^+ + 12e^- = C_4H_8N_2O_3 + 9H_2O$
Aspartate	30.5	$4HCO_3^- + NH_4^+ + 14H^+ + 12e^- = C_4H_7NO_4^- + 8H_2O$
Cysteine	10.5	$3HCO_3^- + NH_4^+ + H_2S + 12H^+ + 10e^- = C_3H_7NO_2S + 7H_2O$
Glutamate	40.8	$5HCO_3^- + NH_4^+ + 21H^+ + 18e^- = C_5H_8NO_4^- + 11H_2O$
Glutamine	36.5	$5HCO_3^- + 2NH_4^+ + 21H^+ + 18e^- = C_5H_{10}N_2O_3 + 12H_2O$
Glycine	43.7	$2HCO_3^- + NH_4^+ + 7H^+ + 6e^- = C_2H_5NO_2 + 4H_2O$
Histidine	14.0	$6HCO_3^- + 3\ NH_4^+ + 23H^+ + 20e^- = C_6H_9N_3O_2 + 16H_2O$
Isoleucine	36.2	$6HCO_3^- + NH_4^+ + 35H^+ + 30e^- = C_6H_{13}NO_2 + 16H_2O$
Leucine	56.1	$6HCO_3^- + NH_4^+ + 35H^+ + 30e^- = C_6H_{13}NO_2 + 16H_2O$
Lysine$^+$	47.7	$6HCO_3^- + 2NH_4^+ + 33H^+ + 28e^- = C_6H_{15}N_2O_2^+ + 16H_2O$
Methionine	21.8	$5HCO_3^- + NH_4^+ + H_2S + 26H^+ + 22e^- = C_5H_{11}NO_2S + 13H_2O$
Phenylalanine	29.1	$9HCO_3^- + NH_4^+ + 48H^+ + 40e^- = C_9H_{11}NO_2 + 25H_2O$
Proline	24.2	$5HCO_3^- + NH_4^+ + 26H^+ + 22e^- = C_5H_9NO_2 + 13H_2O$
Serine	21.5	$3HCO_3^- + NH_4^+ + 12H^+ + 10e^- = C_3H_7NO_3 + 6H_2O$
Threonine	28.7	$4HCO_3^- + NH_4^+ + 19H^+ + 16e^- = C_4H_9N_4O_3 + 9H_2O$
Tryptophan	11.0	$11HCO_3^- + 2NH_4^+ + 55H^+ + 46e^- = C_{11}H_{12}N_2O_2 + 31H_2O$
Tyrosine	23.7	$9HCO_3^- + NH_4^+ + 46H^+ + 38e^- = C_9H_{11}NO_3 + 24H_2O$
Valine	27.1	$5HCO_3^- + NH_4^+ + 28H^+ + 24e^- = C_5H_{11}NO_2 + 13H_2O$
AMP^{2-}	57.3	$10HCO_3^- + 5NH_4^+ + HPO_4^{2-} + 35H^+ + 30e^- = C_{10}H_{12}N_5O_7P^{2-} + 27H_2O$
GMP^{2-}	40.7	$10HCO_3^- + 5NH_4^+ + HPO_4^{2-} + 33H^+ + 28e^- = C_{10}H_{12}N_5O_8P^{2-} + 26H_2O$
CMP^{2-}	73.7	$9HCO_3^- + 3NH_4^+ + HPO_4^{2-} + 36H^+ + 30e^- = C_9H_{12}N_3O_8P^{2-} + 23H_2O$
UMP^{2-}	44.1	$9HCO_3^- + 2NH_4^+ + HPO_4^{2-} + 37H^+ + 30e^- = C_9H_{11}N_2O_9P^{2-} + 22H_2O$
dAMP^{2-}	8.2	$10HCO_3^- + 5NH_4^+ + HPO_4^{2-} + 37H^+ + 32e^- = C_{10}H_{12}N_5O_6P^{2-} + 28H_2O$

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table IIa. (continued)

Compound	*Amt.*[a] *[mg (g cells)$^{-1}$]*	*Biomass Synthesis Reaction*
dGMP^{2-}	7.8	$10HCO_3^- + 5NH_4^+ + HPO_4^{2-} + 35H^+ + 30e^- = C_{10}H_{12}N_5O_7P^{2-} + 27H_2O$
dCMP^{2-}	8.8	$9HCO_3^- + 3NH_4^+ + HPO_4^{2-} + 38H^+ + 32e^- = C_9H_{12}N_3O_7P^{2-} + 24H_2O$
dTMP^{2-}	7.9	$10HCO_3^- + 2NH_4^+ + HPO_4^{2-} + 46H^+ + 38e^- = C_{10}H_{13}N_2O_8P^{2-} + 26H_2O$
Palmitate	28.5	$16HCO_3^- + 107H^+ + 92e^- = C_{16}H_{31}O_2^- + 46H_2O$
Oleate	17.5	$18HCO_3^- + 119H^+ + 102e^- = C_{18}H_{33}O_2^- + 52H_2O$
Palmitoleate	21.7	$16HCO_3^- + 105H^+ + 90e^- = C_{16}H_{29}O_2^- + 46H_2O$
Myristate	3.8	$14HCO_3^- + 93H^+ + 80e^- = C_{14}H_{27}O_2^- + 40H_2O$
β-Hydroxymyristate	8.2	$14HCO_3^- + 91H^+ + 78e^- = C_{14}H_{27}O_3^- + 39H_2O$
Glycerol	15.3	$3HCO_3^- + 17H^+ + 14e^- = C_3H_8O_3 + 6H_2O$
Glucose	4.5	$6HCO_3^- + 30H^+ + 24e^- = C_6H_{12}O_6 + 12H_2O$
Heptose	5.3	$7HCO_3^- + 35H^+ + 28e^- = C_7H_{14}O_7 + 14H_2O$
Galactose	1.5	$6HCO_3^- + 30H^+ + 24e^- = C_6H_{12}O_6 + 12H_2O$
Rhamnose	1.4	$6HCO_3^- + 32H^+ + 26e^- = C_6H_{12}O_5 + 13H_2O$
Glucoseamine	3.0	$6HCO_3^- + NH_4^+ + 29H^+ + 24e^- = C_6H_{13}NO_5 + 13H_2O$
N-acetylglucoseamine	8.0	$8HCO_3^- + NH_4^+ + 39H^+ + 32e^- = C_8H_{15}NO_6 + 18H_2O$
N-acetylmuramic acid	8.1	$11HCO_3^- + NH_4^+ + 54H^+ + 44e^- = C_{11}H_{19}NO_8 + 25H_2O$
Ethanolamine	8.0	2HCO3- + NH4+ + 11H+ + 10e- = C2H7NO + 5H2O
Diaminopimelic acid	5.3	$7HCO_3^- + 2NH_4^+ + 33H^+ + 28e^- = C_7H_{14}N_2O_4 + 17H_2O$
Putrescine	3.0	$4HCO_3^- + 2NH_4^+ + 24H^+ + 22e^- = C_4H_{12}N_2 + 12H_2O$
Spermidine	1.0	$7HCO_3^- + 3NH_4^+ + 42H^+ + 38e^- = C_7H_{19}N_3 + 21H_2O$

[a] Amount of each monomer per gram dry weight of cells, based on the composition of *Escherichia coli* given by Battley (*69*).

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table IIb. Values of $\Delta G_r°$ at 250 bar and Several Temperatures for the Synthesis Reactions Given in Table IIa

	$\Delta G_r°$ (kJ mol^{-1})					
	25°C	*50°C*	*75°C*	*100°C*	*125°C*	*150°C*
Alanine	-189.2	-196.1	-204.2	-213.6	-224.2	-235.9
Arginine$^+$	-243.3	-255.5	-270.4	-287.8	-307.6	-329.8
Asparagine	-150.4	-157.6	-166.3	-176.4	-188.0	-200.9
Aspartate	-167.4	-174.0	-181.9	-191.0	-201.2	-212.6
Cysteine	-119.9	-122.4	-126.2	-131.3	-137.6	-145.1
Glutamate	-291.0	-300.8	-312.2	-325.2	-339.8	-355.9
Glutamine	-277.2	-288.0	-300.8	-315.6	-332.3	-350.9
Glycine	-74.7	-78.5	-82.9	-87.9	-93.6	-99.8
Histidine	-232.9	-246.1	-262.0	-280.4	-301.1	-324.3
Isoleucine	-531.2	-548.6	-569.1	-592.5	-618.7	-647.6
Leucine	-540.4	-557.9	-578.5	-602.0	-628.4	-657.5
Lysine$^+$	-497.7	-513.5	-532.0	-553.2	-576.9	-603.1
Methionine	-533.3	-543.7	-556.5	-571.6	-588.9	-608.4
Phenylalanine	-766.2	-793.4	-825.2	-861.2	-901.3	-945.5
Proline	-372.6	-386.8	-403.2	-421.8	-442.6	-465.4
Serine	-99.4	-105.3	-112.3	-120.4	-129.5	-139.7
Threonine	-205.7	-214.5	-224.9	-237.0	-250.7	-266.0
Tryptophan	-839.8	-870.7	-906.9	-948.2	-994.4	-1045.3
Tyrosine	-706.6	-732.3	-762.3	-796.3	-834.3	-876.2
Valine	-421.7	-435.5	-451.9	-470.6	-491.5	-514.8
AMP^{2-}	-75.9	-93.6	-115.7	-142.1	-172.7	-207.3
GMP^{2-}	-86.3	-102.2	-122.4	-146.7	-175.1	-207.5
CMP^{2-}	-223.8	-241.7	-263.6	-289.1	-318.2	-350.9

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table IIb (continued)

	$\Delta G_r°$ (kJ mol^{-1})					
	25 °C	*50 °C*	*75 °C*	*100 °C*	*125 °C*	*150 °C*
UMP^{2-}	-280.9	-299.6	-322.2	-348.6	-378.5	-411.9
dAMP^{2-}	-164.9	-184.1	-207.9	-236.0	-268.3	-304.8
dGMP^{2-}	-180.7	-197.4	-218.5	-243.8	-273.2	-306.6
dCMP^{2-}	-319.0	-339.1	-363.3	-391.2	-422.7	-457.9
dTMP^{2-}	-507.4	-531.8	-560.8	-594.0	-631.4	-672.8
Palmitate	-1743.1	-1800.4	-1867.1	-1942.3	-2025.8	-2117.4
Oleate	-1887.2	-1952.6	-2028.3	-2113.7	-2208.2	-2311.7
Palmitoleate	-1659.0	-1716.9	-1783.9	-1859.5	-1943.2	-2035.0
Myristate	-1514.9	-1564.7	-1622.6	-1688.0	-1760.6	-1840.1
β-Hydroxymyristate	-1427.5	-1476.2	-1532.8	-1596.9	-1668.1	-1746.1
Glycerol	-148.1	-157.9	-168.9	-181.1	-194.4	-208.9
Glucose	-235.9	-250.5	-267.7	-287.4	-309.6	-334.3
Heptose	-285.4	-303.0	-323.7	-347.3	-373.8	-403.1
Galactose	-229.0	-243.6	-260.8	-280.7	-303.0	-328.0
Rhamnose	-308.1	-323.7	-342.4	-363.8	-388.0	-414.9
Glucoseamine	-190.4	-203.9	-220.0	-238.7	-259.8	-283.5
N-acetylglucoseamine	-406.1	-426.1	-449.7	-476.7	-507.1	-540.8
N-acetylmuramic acid	-657.3	-686.0	-719.6	-757.6	-800.2	-847.2
Ethanolamine	-68.8	-74.7	-81.6	-89.3	-97.9	-107.4
Diaminopimelic acid	-402.7	-422.2	-444.4	-469.1	-496.4	-526.2
Putrescine	-257.5	-270.1	-284.8	-301.5	-320.2	-340.8
Spermidine	-482.1	-503.8	-529.1	-557.9	-590.0	-625.3

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table III. Values of $\Delta G_i°$ at 250 bar and Several Temperatures for Compounds Used in the Energy Calculations[a]

Compound	*$\Delta G_i°$ (kJ mol^{-1})*					
	25°C	*50°C*	*75°C*	*100°C*	*125°C*	*150°C*
Inorganic Compounds						
H_2O^a	-236.73	-238.55	-240.52	-242.62	-244.84	-247.19
H^{+a}	0.00	0.00	0.00	0.00	0.00	0.00
e^{-a}	0.00	0.00	0.00	0.00	0.00	0.00
HCO_3^{-b}	-586.33	-588.74	-591.12	-593.48	-595.81	-598.07
NH_4^{+b}	-79.00	-81.84	-84.82	-87.91	-91.11	-94.42
$H_2(aq)^c$	18.35	16.77	14.90	12.77	10.40	7.82
$H_2S(aq)^c$	-33.56	-38.74	-43.99	-49.30	-54.67	-60.10
HPO_4^{2-b}	-1089.01	-1087.96	-1086.55	-1084.80	-1082.72	-1080.28
NO_3^{-b}	-110.19	-113.76	-117.26	-120.71	-124.10	-127.41
NO_2^{-b}	-31.60	-34.57	-37.42	-40.16	-42.79	-45.30
SO_4^{2-b}	-744.12	-744.31	-744.13	-743.62	-742.79	-741.60
Amino Acids[d]						
Alanine	-370.07	-374.26	-378.79	-383.64	-388.80	-394.25
Arginine$^+$	-289.52	-298.46	-308.11	-318.45	-329.42	-340.97
Asparagine	-523.11	-529.27	-535.77	-542.64	-549.84	-557.37
Aspartate	-697.81	-702.35	-707.04	-711.87	-716.82	-721.86
Cysteine	-334.26	-339.28	-344.74	-350.61	-356.87	-363.50
Glutamate	-697.60	-702.20	-706.93	-711.77	-716.68	-721.62
Glutamine	-526.02	-532.70	-539.84	-547.43	-555.45	-563.88
Glycine	-379.45	-383.59	-387.87	-392.31	-396.91	-401.64
Histidine	-200.11	-207.21	-214.88	-223.09	-231.81	-241.02
Isoleucine	-340.40	-345.96	-352.30	-359.37	-367.12	-375.52

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table III (continued)

Compound	$\Delta G_i°$ *(kJ mol^{-1})*					
	25 ℃	*50 ℃*	*75 ℃*	*100 ℃*	*125 ℃*	*150 ℃*
Leucine	-349.60	-355.24	-361.70	-368.91	-376.82	-385.40
Lysine^{+}	-385.96	-392.70	-400.06	-408.00	-416.48	-425.44
Methionine	-499.95	-506.73	-514.15	-522.18	-530.78	-539.91
Phenylalanine	-203.77	-210.04	-217.14	-225.03	-233.64	-242.94
Proline	-305.71	-311.09	-316.90	-323.13	-329.76	-336.78
Serine	-517.01	-522.05	-527.39	-533.05	-539.00	-545.23
Threonine	-499.43	-504.25	-509.53	-515.27	-521.44	-528.02
Tryptophan	-108.63	-115.26	-122.79	-131.16	-140.31	-150.21
Tyrosine	-380.98	-387.47	-394.73	-402.74	-411.45	-420.82
Valine	-354.75	-359.82	-365.53	-371.85	-378.73	-386.15
Nucleotides[e]						
AMP^{2-}	-1031.37	-1037.12	-1043.55	-1050.62	-1058.26	-1066.40
GMP^{2-}	-1278.54	-1284.27	-1290.71	-1297.83	-1305.55	-1313.81
CMP^{2-}	-1381.93	-1387.10	-1392.72	-1398.76	-1405.16	-1411.86
UMP^{2-}	-1596.69	-1601.67	-1607.10	-1612.94	-1619.14	-1625.64
dAMP^{2-}	-883.67	-889.15	-895.24	-901.89	-909.07	-916.68
dGMP^{2-}	-1136.20	-1140.92	-1146.28	-1152.25	-1158.75	-1165.72
dCMP^{2-}	-1240.35	-1245.93	-1251.88	-1258.19	-1264.81	-1271.66
dTMP^{2-}	-1462.66	-1468.43	-1474.68	-1481.38	-1488.47	-1495.87
Fatty Acids						
Palmitate[f]	-234.60	-246.70	-261.13	-277.61	-295.95	-315.99
Oleate[g]	-130.92	-144.99	-161.52	-180.23	-200.90	-223.39
Palmitoleate[h]	-150.44	-163.12	-177.98	-194.79	-213.36	-233.59
Myristate[i]	-254.16	-264.85	-277.55	-292.03	-308.12	-325.69

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table III (continued)

Compound	ΔG_i° (*kJ mol*$^{-1}$)					
	25 °C	*50 °C*	*75 °C*	*100 °C*	*125 °C*	*150 °C*
β-Hydroxymyristate[j]	-403.43	-414.84	-428.30	-443.58	-460.49	-478.92
Saccharides						
Glycerol[k]	-486.65	-492.77	-499.15	-505.82	-512.78	-520.01
Glucose[l]	-913.10	-920.24	-928.16	-936.85	-946.29	-956.47
Heptose[m]	-1075.43	-1084.43	-1094.32	-1105.07	-1116.64	-1129.01
Galactose[l]	-906.17	-913.32	-921.31	-930.13	-939.76	-950.18
Rhamnose[n]	-748.55	-754.95	-762.34	-770.66	-779.85	-789.91
Glucoseamine[o]	-709.83	-716.95	-724.82	-733.45	-742.82	-752.92
N-acetylglucoseamine[p]	-914.50	-923.85	-934.15	-945.38	-957.50	-970.51
N-acetylmuramic acid[q]	-1267.59	-1280.12	-1293.74	-1308.41	-1324.11	-1340.83

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table III (continued)

Compound	*$\Delta G_i{}^o$ (kJ mol^{-1})*					
	25°C	*50°C*	*75°C*	*100°C*	*125°C*	*150°C*
Amines						
Ethanolamine[r]	-136.77	-141.26	-146.03	-151.09	-156.44	-162.08
Diaminopimelic acid[s]	-640.55	-651.63	-663.04	-674.81	-686.95	-699.44
Putrescine[t]	79.96	73.92	67.31	60.15	52.46	44.26
Spermidine[u]	148.02	139.15	129.47	118.99	107.76	95.80

[a] Calculated with SUPCRT (*61*). [b] Shock and Helgeson (1988) (*63*). [c] Shock et al. (1989) (*64*). [d] Dick et al. (2006) (*76*). [e] LaRowe and Helgeson (2006) (*77*). [f] Properties calculated with Palmitate = Propanoate + 13($-CH_2-$); values of Propanoate from Shock (1995) (*78*) and $-CH_2-$ from Amend and Helgeson (1997) (*79*). [g] Calculated with Oleate = Palmitate + 2($-CH_2-$) + (Octene – Octane); values of Octene and Octane from Shock and Helgeson (1990) (*80*). [h] Calculated with Palmitoleate = Palmitate + (Octene – Octane). [i] Calculated with Myristate = Propanoate + 11($-CH_2-$). [j] Calculated with β-Hydroxymyristate = Myristate + (Hydroxyoctanoate – Octanoate); values of Octanoate and Hydroxyoctanoate from Shock (1995) (*78*). [k] ΔG_f° from Thauer et al. (1977) (*81*); other properties calculated with Glycerol = 3($-CH_2OH$) + ($-CH_2-$) – ($-CH_3$) with groups from Amend and Helgeson (1997) (*79*). [l] Amend and Plyasunov (2001) (*82*). [m] Calculated with Heptose = C_6-aldose + (C_6-aldose – C_5-aldose); values of C_6- and C_5-aldoses are average values in Amend and Plyasunov (2001) (*82*). [n] Calculated with Rhamnose = Glucose + (Hexane – Hexanol); values of Glucose from Amend and Plyasunov (2001) (*82*) Hexane and Hexanol from Amend and Helgeson (1997) (*79*). [o] Calculated with Glucoseamine = Glucose + (Hexanamine – Hexanol); values of Hexanamine and Hexanol from Amend and Helgeson (1997) (*79*). [p] Calculated with N-acetylglucoseamine = Glucose + 2($-CH_2-$) + ($-CONH_2$) – ($-CH_2OH$) with groups from Amend and Helgeson (1997) (*79*). [q] Calculated with N-acetylmuramic acid = N-acetylglucoseamine + ($-COOH$) + ($-CHCH_3$) + ($-CH_2-$) – ($-CH_3$) with groups from Amend and Helgeson (1997) (*79*). [r] Calculated with Ethanolamine = Ethanamine + (Ethanol – Ethane); values of Ethanamine, Ethanol, and Ethane from Amend and Helgeson (1997) (*79*). [s] Calculated with Diaminopimelic acid = Pimelic acid + 2($-CH_2NH_2$) – 2($-CH_3$); values of Pimelic acid from Shock (1995) (*78*) and groups from Amend and Helgeson (1997) (*79*). [t] Calculated with Putrescine = 2($-CH_2-$) + 2($-CH_2NH_2$). [u] Calculated with Spermidine = Putrescine + 3($-CH_2-$) + ($-CH_2NH_2$) – ($-CH_3$).

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table IV. Values of ΔG_r (in kJ mol^{-1}) at 250 bar and Several Temperatures for the Synthesis Reactions Given in Table IIa

	Amount		*ΔG_r (kJ mol^{-1})*				
	mg (g cells)$^{-1}$	*μmol (g cells)$^{-1}$*	*25 °C*	*50 °C*	*75 °C*	*100 °C*	*125 °C*
Amino Acids							
Alanine	48.4	543	0	-114	-105	-89	-55
Arginine$^+$	49.0	281	240	46	76	118	191
Asparagine	30.3	229	96	-12	3	26	69
Aspartate	30.5	229	33	-77	-66	-45	-5
Cysteine	10.5	87	92	-3	11	31	71
Glutamate	40.8	278	10	-158	-142	-113	-58
Glutamine	36.5	250	69	-96	-77	-46	11
Glycine	43.7	582	14	-43	-38	-29	-11
Histidine	14.0	90	188	8	32	68	133
Isoleucine	36.2	276	-42	-328	-307	-268	-190
Leucine	56.1	428	-51	-337	-316	-278	-200
Lysine$^+$	47.7	326	20	-241	-216	-173	-91
Methionine	21.8	146	-122	-332	-311	-276	-208
Phenylalanine	29.1	176	-45	-426	-397	-342	-227
Proline	24.2	210	0	-210	-195	-165	-104
Serine	21.5	205	73	-21	-12	4	36
Threonine	28.7	241	66	-84	-71	-47	2
Tryptophan	11.0	54	39	-393	-354	-284	-146
Tyrosine	23.7	131	-2	-363	-333	-278	-163
Valine	27.1	402	-32	-261	-244	-213	-150
Nucleotides							
AMP^{2-}	57.3	165	700	445	495	565	683

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table IV. (continued)

	Amount		ΔG_r (kJ mol^{-1})				
	mg (g cells)$^{-1}$	*μmol (g cells)$^{-1}$*	*25 °C*	*50 °C*	*75 °C*	*100 °C*	*125 °C*
GMP^{2-}	40.7	126	672	438	490	559	677
CMP^{2-}	73.7	203	486	221	263	326	439
UMP^{2-}	44.1	136	421	151	191	252	367
$dAMP^{2-}$	8.2	25	628	353	402	472	591
$dGMP^{2-}$	7.8	25	595	341	393	463	583
$dCMP^{2-}$	8.8	25	408	122	163	225	337
$dTMP^{2-}$	7.9	25	311	-37	6	74	204
Fatty Acids							
Palmitate	28.5	111	-281	-1169	-1116	-1010	-791
Oleate	17.5	62	-242	-1227	-1168	-1050	-804
Palmitolate	21.7	85	-214	-1084	-1032	-928	-712
Myristate	3.8	17	-253	-1026	-981	-889	-699
β-Hydroxymyristate	8.2	34	-183	-935	-890	-799	-610

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table IV. (continued)

	Amount		ΔG_r *(kJ mol^{-1})*				
	mg (g cells)$^{-1}$	*µmol (g cells)$^{-1}$*	*25 °C*	*50 °C*	*75 °C*	*100 °C*	*125 °C*
Saccharides							
Glycerol	15.3	161	50	-90	-84	-69	-33
Glucose	4.5	25	194	-36	-16	20	98
Heptose	5.3	25	227	-40	-17	25	117
Galactose	1.5	8.4	201	-29	-9	27	104
Rhamnose	1.4	8.4	139	-111	-92	-56	23
Glucoseamine	3.0	17	248	23	45	83	159
N-acetylglucoseamine	8.0	36	198	-103	-74	-24	78
N-acetylmuramic acid	8.1	28	196	-219	-179	-109	34
Amines							
Ethanolamine	8.0	131	54	-43	-39	-29	-9
Diaminopimelic acid	5.3	28	127	-136	-111	-68	16
Putrescine	3.0	34	74	-134	-120	-94	-47
Spermidine	1.0	7	142	-215	-187	-138	-50

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table V. Values of ΔG_r [in J (cells)$^{-1}$] at 250 bar and Several Temperatures for the Synthesis Reactions Given in Table IIa

	Amount		*ΔG_r [J (g cells)$^{-1}$]*				
	mg (g cells)$^{-1}$	*µmol (g cells)$^{-1}$*	*25 ℃*	*50 ℃*	*75 ℃*	*100 ℃*	*125 ℃*
Amino Acids							
Alanine	48.4	543	-0.1	-61.9	-57.1	-48.2	-29.9
Arginine$^+$	49.0	281	67.6	13.0	21.3	33.2	53.7
Asparagine	30.3	229	22.0	-2.6	0.8	6.0	15.8
Aspartate	30.5	229	7.6	-17.7	-15.0	-10.4	-1.2
Cysteine	10.5	87	8.0	-0.2	0.9	2.7	6.2
Glutamate	40.8	278	2.7	-43.9	-39.3	-31.4	-16.0
Glutamine	36.5	250	17.3	-23.9	-19.2	-11.6	2.7
Glycine	43.7	582	8.3	-24.8	-22.1	-17.1	-6.3
Histidine	14.0	90	16.9	0.7	2.8	6.0	12.0
Isoleucine	36.2	276	-11.5	-90.5	-84.7	-74.0	-52.5
Leucine	56.1	428	-21.8	-144.3	-135.4	-118.9	-85.6
Lysine$^+$	47.7	326	6.6	-78.7	-70.3	-56.3	-29.6
Methionine	21.8	146	-17.8	-48.4	-45.4	-40.3	-30.3
Phenylalanine	29.1	176	-7.9	-75.0	-69.8	-60.1	-39.9
Proline	24.2	210	-0.1	-44.2	-40.9	-34.7	-21.8
Serine	21.5	205	14.9	-4.4	-2.5	0.7	7.4
Threonine	28.7	241	16.0	-20.3	-17.0	-11.2	0.4
Tryptophan	11.0	54	2.1	-21.2	-19.1	-15.4	-7.9
Tyrosine	23.7	131	-0.3	-47.6	-43.6	-36.4	-21.4
Valine	27.1	402	-13.0	-104.9	-98.1	-85.5	-60.2
Total Amino Acids	**631**	**5160**	**117**	**-841**	**-754**	**-603**	**-305**

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table V. (continued)

	Amount		ΔG_r [J (g cells)$^{-1}$]				
	mg (g cells)$^{-1}$	*µmol (g cells)$^{-1}$*	*25 °C*	*50 °C*	*75 °C*	*100 °C*	*125 °C*
Nucleotides							
AMP^{2-}	57.3	165	115.5	73.4	81.7	93.2	112.7
GMP^{2-}	40.7	126	84.7	55.2	61.7	70.5	85.4
CMP^{2-}	73.7	203	98.7	44.9	53.5	66.2	89.0
UMP^{2-}	44.1	136	57.2	20.6	26.0	34.3	49.9
$dAMP^{2-}$	8.2	25	15.5	8.7	9.9	11.6	14.5
$dGMP^{2-}$	7.8	25	15.1	8.7	10.0	11.8	14.8
$dCMP^{2-}$	8.8	25	10.4	3.1	4.1	5.7	8.6
$dTMP^{2-}$	7.9	25	7.7	-0.9	0.1	1.8	5.0
Total Nucleotides	**249**	**730**	**405**	**214**	**247**	**295**	**380**
Fatty Acids							
Palmitate	28.5	111	-31.2	-129.7	-123.9	-112.1	-87.8
Oleate	17.5	62	-15.0	-76.0	-72.3	-65.0	-49.8
Palmitolate	21.7	85	-18.2	-92.3	-87.9	-79.0	-60.6
Myristate	3.8	17	-4.3	-17.2	-16.5	-14.9	-11.7
β-Hydroxymyristate	8.2	34	-6.1	-31.4	-29.9	-26.8	-20.5
Total Fatty Acids	**80**	**308**	**-75**	**-347**	**-330**	**-298**	**-230**

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Table V. (continued)

	Amount		*ΔG_r [J (g cells)$^{-1}$]*				
	mg (g cells)$^{-1}$	*μmol (g cells)$^{-1}$*	*25 ℃*	*50 ℃*	*75 ℃*	*100 ℃*	*125 ℃*
Saccharides							
Glycerol	15.3	161	8.0	-14.4	-13.6	-11.2	-5.3
Glucose	4.5	25	4.9	-0.9	-0.4	0.5	2.5
Heptose	5.3	25	5.8	-1.0	-0.4	0.6	3.0
Galactose	1.5	8.4	1.7	-0.2	-0.1	0.2	0.9
Rhamnose	1.4	8.4	1.2	-0.9	-0.8	-0.5	0.2
Glucoseamine	3.0	17	4.2	0.4	0.8	1.4	2.7
N-acetylglucoseamine	8.0	36	7.1	-3.7	-2.7	-0.9	2.8
N-acetylmuramic acid	8.1	28	5.4	-6.0	-4.9	-3.0	0.9
Total Saccharides	**47**	**309**	**38**	**-27**	**-22**	**-13**	**8**
Amines							
Ethanolamine	8.0	131	7.1	-5.6	-5.0	-3.8	-1.1
Diaminopimelic acid	5.3	28	3.5	-3.7	-3.1	-1.9	0.5
Putrescine	3.0	34	2.5	-4.6	-4.1	-3.2	-1.6
Spermidine	1.0	7	1.0	-1.5	-1.3	-1.0	-0.3
Total Amines	**17**	**199**	**14**	**-15**	**-13**	**-10**	**-3**
Total Cell Biomass	**1024**[a]		**500**	**-1016**	**-873**	**-628**	**-150**

[a] The total exceeds 1 g, because some water is lost by dehydration during polymerization.

References

1. Mojzsis, S. J.; Arrhenius, G.; McKeegan, K. D.; Harrison, T. M.; Nutman, A. P.; Friend, C. R. L. *Nature* **1996,** *384,* 55–59.
2. Rosing, M. T. *Science* **1999,** *283,* 674–676.
3. Schopf, J. W.; Kudryavtsev, A. B.; Agresti, D. G.; Wdowiak, T. J.; Czaja, A. D. *Nature* **2002,** *416,* 73–76.
4. Schopf, J. W.; Packer, B. M. *Science* **1987,** *237,* 70–73.
5. Brasier, M. D.; Green, O. R.; Jephcoat, A. P.; Kleppe, A. K.; Van Kranendonk, M. J.; Lindsay, J. F.; Steele, A.; Grassineau, N. V. *Nature* **2002,** *416,* 76–81.
6. Pasteris, J. D.; Wopenka, B. *Nature* **2002,** *420,* 476–477.
7. Rye, R.; Holland, H. D. *Am. J. Sci.* **1998,** *298,* 621–672.
8. Brocks, J. J.; Logan, G. A.; Buick, R.; Summons, R. E. *Science* **1999,** *285,* 1033–1036.
9. Hedges, S. B.; Chen, H.; Kumar, S.; Wang, D. Y.-C.; Thompson, A. S.; Watanabe, H. *BMC Evol. Biol.* **2001,** *1,* 10 pages.
10. Xiong, J.; Bauer, C. E., *Annual Reviews of Plant Biology* **2002,** 53, 503–521.
11. Xiong, J.; Fischer, W. M.; Inoue, K.; Nakahara, M.; Bauer, C. E. *Science* **2000,** *289,* 1724–1730.
12. Blankenship, R. E. *Molecular mechanisms of photosynthesis;* Blackwell Science Ltd.: Oxford, UK, 2002; p. 321.
13. Raymond, J.; Zhaxybayeva, O.; Gogarten, J. P.; Gerdes, S. Y.; Blankenship, R. E. *Science* **2002,** *298,* 1616–1620.
14. Tice, M. M.; Lowe, D. R. *Nature* **2004,** *431,* 549–552.
15. Tice, M. M.; Lowe, D. R. *Geology* **2006,** *34,* 37–40.
16. Martin, W.; Russell, M. J. *Phil. Trans. Royal Soc. London B* **2007,** *362,* 1887-1925.
17. Mereschkowsky, C. *Biol. Centralbl.* **1910,** *30,* 278–288 & 289–303 & 321–347 & 353–367.
18. Russell, M. J.; Hall, A. J., A hydrothermal source of energy and materials at the origin of life. In this volume.
19. Fuchs, G. In *Autotrophic Bacteria;* Schlegel, H. G.; Bowien, B., Eds.; Science Tech.: Madison, WI, 1989; pp 365–382.
20. Shock, E. L.; McCollom, T.; Schulte, M. D. *Origins Life Evol. Biosphere* **1995,** *25,* 141–159.
21. McCollom, T. M.; Shock, E. L. *Geochim. Cosmochim. Acta* **1997,** *61,* 4375–4391.
22. Amend, J. P.; Shock, E. L. *Science* **1998,** *281,* 1659–1662.
23. Shock, E. L.; Schulte, M. D. *J. Geophys. Res.* **1998,** *103,* 28513–28527.
24. Shock, E. L. In *Evolution of Hydrothermal Ecosystems on Earth (and Mars?);* Bock, G. R.; Goode, J. A., Eds.; John Wiley & Sons: Chichester, New York, 1996; pp 40–60.
25. Shock, E. L.; Amend, J. P.; Zolotov, M. Y. In *Origin of the Earth and Moon;* Canup, R.; Righter, K., Eds.; The University of Arizona Press: Tucson, AZ, 2000; pp 527–543.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

26. Amend, J. P.; Rogers, K. L.; Shock, E. L.; Gurrieri, S.; Inguaggiato, S. *Geobiology* **2003,** *1,* 37–58.
27. Pace, N. R. *Cell* **1991,** *65,* 531–533.
28. Pace, N. R. *Science* **1997,** *276,* 734–740.
29. Corliss, J. B.; Baross, J. A.; Hoffman, S. E. *Oceanolog. Acta* **1981,** no. Sp, 59–69.
30. Baross, J. A.; Hoffman, S. E. *Origins of Life* **1985,** *15,* 327–345.
31. Holm, N. G. In *Marine Hydrothermal Systems and the Origin of Life;* Holm, N. G.; Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1992; pp 5-14.
32. Russell, M. J.; Daniel, R. M.; Hall, A. J.; Sherringham, J. *Journal of Molecular Evolution* **1994,** *39,* 231–243.
33. Russell, M. J.; Hall, A. J. *J. Geol. Soc. (London)* **1997,** *154,* 377–402.
34. Sleep, N. H.; Zahnle, K. J.; Kasting, J. F.; Morowitz, H. J. *Nature* **1989,** *342,* 139–142.
35. Chyba, C. F. *Geochim. Cosmochim. Acta* **1993,** *57,* 3351–3358.
36. Zahnle, K.; Arndt, N.; Cockell, C.; Halliday, A.; Nisbet, E.; Selsis, F.; Sleep, N. H. *Space Sci. Rev.* **2007,** *129,* 35-78.
37. Bada, J. L. *Earth Planet. Sci. Lett.* **2004,** *226,* 1–15.
38. Shock, E. L. *J. Geophys. Res.* **1997,** *102,* 23687–23694.
39. Jakosky, B. M.; Shock, E. L. *J. Geophys. Res.* **1998,** *103,* 19359–19364.
40. McCollom, T. M. *J. Geophys. Res.* **1999,** *104,* 30729–30742.
41. Shock, E. L. *Origins Life Evol. Biosphere* **1990,** *20,* 331–367.
42. Amend, J. P.; Shock, E. L. In *Perspectives in amino acid and protein geochemistry;* Goodfriend, G.; Collins, M. J.; Fogel, M. L.; Macko, S. A.; Wehmiller, J. F., Eds.; Oxford University Press: New York, 2000; pp 23–40.
43. Kasting, J. F. *Science* **1993,** *259,* 920–926.
44. Catling, D. C.; Claire, M. W. *Earth Planet. Sci. Lett.* **2005,** *237,* 1–20.
45. Kasting, J. F. In *Organic Geochemistry;* Engel, M. H.; Macko, S. A., Eds.; Plenum Press: New York, NY, 1993; pp 611–624.
46. Pavlov, A. A.; Kasting, J. F.; Brown, L. L.; Rages, K. A.; Freedman, R. *J. Geophys. Res.* **2000,** *105,* 11981-11990.
47. Kasting, J. F.; Howard, M. T. *Phil. Trans. Royal Soc. London B* **2006,** *361,* 1733–1742.
48. Knauth, L. P.; Lowe, D. R. *Geol. Soc. Am. Bull.* **2003,** *115,* 566–580.
49. Robert, F.; Chaussidon, M. *Nature* **2006,** *443,* 969–972.
50. Tian, F.; Toon, O. B.; Pavlov, A. A.; De Sterck, H. *Science* **2005,** *308,* 1014–1017.
51. Pavlov, A. A.; Kasting, J. F. *Astrobiology* **2002,** *2,* 27–41.
52. Habicht, K. S.; Gade, M.; Thamdrup, B.; Berg, P.; Canfield, D. E. *Science* **2002,** *298,* 2372–2374.
53. Russell, M. J.; Hall, A. J. In *Evolution of Early Earth's Atmosphere, Hydrosphere, and Biosphere -- Constraints from Ore Deposits;* Kesler, S. E.; Ohmoto, H., Eds.; Geological Society of America: Boulder, CO, 2006; GSA Memoir 198, pp 1–32.
54. de Wit, M. J.; Hart, R. A. *Lithos* **1993,** *30,* 309-335.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

55. Condie, K. C. In *Archean Crustal Evolution;* Condie, K. C., Ed.; Elsevier: Amsterdam, 1994; 85-120.
56. Kelley, D. S.; Karson, J. A.; Blackman, D. K.; Frueh-Green, G. L.; Butterfield, D. A.; Lilley, M. D.; Olson, E. J.; Schrenk, M. O.; Roe, K. K.; Lebon, G. T.; Rivizzigno, P.; Party, A.-S. *Nature* **2001,** *412,* 145–149.
57. Kelley, D. S.; Karson, J. A.; Frueh-Green, G. L.; Yoerger, D. R.; Shank, T. M.; Butterfield, D. A.; Hayes, J. M.; Schrenk, M. O.; Olson, E. J.; Proskurowski, G.; Jakuba, M.; Bradley, A.; Larson, B.; Ludwig, K.; Glickson, D.; Buckman, K.; Bradley, A. S.; Brazelton, W. J.; Roe, K.; Elend, M. J.; Delacour, A.; Bernasconi, S. M.; Lilley, M. D.; Baross, J. A.; Summons, R. E.; Sylva, S. P. *Science* **2005,** *307,* 1428–1434.
58. Allen, D. E.; Seyfried, W. E., Jr. *Geochim. Cosmochim. Acta* **2004,** *68,* 1347–1354.
59. Proskurowski, G.; Lilley, M. D.; Kelley, D. S.; Olson, E. J. *Chem. Geol.* **2006,** *229,* 331–343.
60. Wolery, T. J.; Daveler, S. A. *EQ6, a computer program for reaction path modeling of aqueous geochemical systems: Theoretical manual, user's guide, and related documents*, Lawrence Livermore National Lab: 1992.
61. Johnson, J. W.; Oelkers, E. H.; Helgeson, H. C., SUPCRT92: A software package for calculating the standard molal properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bar and 0 to 1000°C. *Computers and Geosciences* **1992,** *18,* (7), 899–947.
62. Helgeson, H. C.; Delany, J. M.; Nesbitt, W. H.; Bird, D. K. *Am. J. Sci.* **1978,** *278A,* 1–229.
63. Shock, E. L.; Helgeson, H. C. *Geochim. Cosmochim. Acta* **1988,** *52,* 2009–2036.
64. Shock, E. L.; Helgeson, H. C.; Sverjensky, D. A. *Geochim. Cosmochim. Acta* **1989,** *53,* 2157–2183.
65. Shock, E. L.; Sassani, D. C.; Willis, M.; Sverjensky, D. A. *Geochim. Cosmochim. Acta* **1997,** *61,* 907–950.
66. Sverjensky, D. A.; Shock, E. L.; Helgeson, H. C. *Geochim. Cosmochim. Acta* **1997,** *61,* 1359–1412.
67. Morowitz, H. J. *Energy Flow in Biology;* Academic Press, New York, NY, 1968, 179 p.
68. Stouthamer, A. H.; Bettenhaussen, C. *Biochim. Biophys. Acta* **1973,** *301,* 53–70.
69. Battley, E. H. *Biotechnol. Bioeng.* **1991,** *38,* 480–492.
70. Battley, E. H. *Biotechnol. Bioeng.* **1992,** *40,* 280–288.
71. Battley, E. H. *Biotechnol. Bioeng.* **1992,** *39,* 589–595.
72. McCollom, T. M.; Amend, J. P. *Geobiology* **2005,** *3,* 135–144.
73. Delano, J. W. *Origins Life Evol. Biosphere* **2001,** *31,* 311-341.
74. Proskurowski, G.; Lilley, M. D.; Seewald, J. S.; Früh-Green, G. L.; Olson, E. J.; Lupton, J. E.; Sylva, S. P.; Kelley, D. S. *Science* **2008,** *319,* 604-607.
75. Barnes, I.; O'Neil, J. R. *Geol. Soc. Am. Bull.* **1969,** *80,* 1947-1960.
76. Dick, J. M.; LaRowe, D. E.; Helgeson, H. C. *Biogeosciences* **2006,** *3,* 311–336.
77. LaRowe, D. E.; Helgeson, H. C. *Geochim. Cosmochim. Acta* **2006,** *70,* 4680–4724.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

78. Shock, E. L. *Am. J. Sci.* **1995,** *295,* 496–580.
79. Amend, J. P.; Helgeson, H. C. *Geochim. Cosmochim. Acta* **1997,** *61,* 11–46.
80. Shock, E. L.; Helgeson, H. C. *Geochim. Cosmochim. Acta* **1990,** *54,* 915–945.
81. Thauer, R. K.; Jungermann, K.; Decker, K. *Bacteriological Reviews* **1977,** *41,* (1), 100–180.
82. Amend, J. P.; Plyasunov, A. V. *Geochim. Cosmochim. Acta* **2001,** *65,* 3901–3917.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch004

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

Chapter 5

Prebiotic Selection of the AT Base-Pair?

A Physical Organic Approach to Understanding AT Base-Pair Stability Indicates Special Stability (*1, 2*)

Jordan R. Quinn,[1] Steven C. Zimmerman,[1,]* Janet E. Del Bene,[2] and Isaiah Shavitt[1]

[1]Department of Chemistry, University of Illinois, Urbana, IL 61801
[2]Department of Chemistry, Youngstown State University, Youngstown, Ohio 44555

The A·T and G·C base-pairs have been investigated experimentally and computationally to examine whether they contain any special stability relative to other base pairs. An empirically based relationship between overall complex stability ($-\Delta G°$) and various possible component interactions is developed to probe the question in a large set of experimentally determined base-pairs and their analogs. Unexpected stability is observed in the A·T pair, and its origin is examined computationally. The results and their possible relationship to prebiotic chemistry are described.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

It has been proposed that in the prebiotic world RNA preceded proteins and functioned both as a catalyst and an information storage unit (*3-5*). Given the hydrolytic instability of cytosine, there is further speculation that this "RNA world" (*6*) initially had a single base pair (A·U) (*7*), Not everyone agrees that A·U is a likely component of a primordial replication system (*8*), but if it were, an obvious question is why A·U? In experiments that simulate possible early conditions on Earth, it was elegantly shown that adenine can be produced from HCN and base in up to a 0.5% yield (*9*). But the low yield and the complex mechanism for formation of adenine (*10*) makes it likely that many other possible candidates were present (*11, 12*). In considering the chemical etiology of the nucleic acids (*13*, *14*) is it possible that A·U might possess some special stability that, say relative to other doubly hydrogen bonded base-pairs, might give it a competitive advantage?

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

Special Stability in Base Pairs

The molecular-level genetic storage units, A·T/U and G·C base-pairs, are clearly special. For example, it is well documented that a single tautomeric form predominates in each of the four bases, disfavoring mismatches. What is less well known is whether the strength of base pairing is enhanced by interactions in addition to the strong N−H···N and N−H···O hydrogen bonds that are present. With regard to special stability, early measurements of association constants (K_{assoc}) of base-pair analogs in chloroform by Rich and coworkers showed that A·T/U and G·C base-pairs are more stable than mismatched complexes or base dimers (*15-17*). The concept of a special *electronic complementarity* (*15*) was most clearly supported by the observation that the K_{assoc} for the A·U base-pair was significantly higher than the K_{dimer} for A·A and U·U, despite all three pairs containing two hydrogen bonds. Jorgensen attributed the greater stability of the A·T base-pair to differences in the strength of primary hydrogen bonds (acidity/basicity) and to secondary electrostatic interactions between proximal hydrogen bonding sites (*18-20*). Rebek (*21*), Hunter (*22*), and others have suggested a CH···O hydrogen bond between H-8 of adenine and O-2 of thymine. In addition, many theoretical studies in the literature have reported the results of electronic structure calculations on the A·T base pair.

To further examine the question of special stability and potentially quantify it, we took a decidedly physical organic and theoretical approach. The goal is to develop a method of empirically predicting base pair stability with the idea that a linear relationship between predicted and experimental values would allow us to identify examples that deviate from expectations.

Phenomenological Approach

The objective of this phenomenological analysis is to establish a correlation between assignable increments and complexation free energy ($-\Delta G^{o}_{298}$). This approach was first used by Jorgensen (*18-20*) and subsequently expanded by others (*23-28*), however, these models apply to a limited group of base pairs. Our approach uses an expanded dataset and introduces new variables with a goal of identifying any special stability in the natural base pairs. Using a combinatorial, multivariate linear regression analysis (*29*), we correlate experimental $-\Delta G^{o}_{298}$ values with component interactions using equations of the general form:

$$\Delta G^{o}_{298} = \text{constant} + a n_{a} + b n_{b} + \ldots \qquad (1)$$

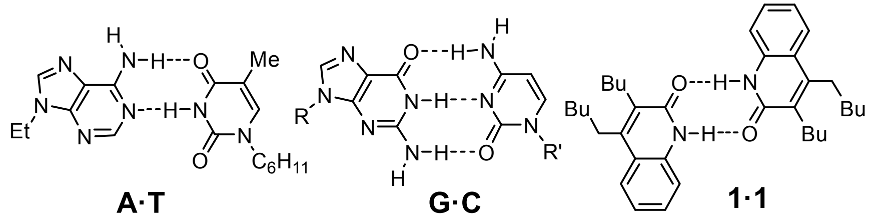
Our dataset contains 256 hydrogen-bonded complexes with association constants (K_{assoc}) in chloroform of at least 1 M^{-1}. Chloroform is a common solvent for studying hydrogen bonding because it weakly competes for hydrogen bonding sites and dissolves many bases. In addition, π-stacking does not interfere with hydrogen bonding measurements in chloroform. It is unlikely that DNA and RNA bases evolved in organic solvent, however, chloroform is the logical choice for this model because hydrogen bonding in chloroform is so widely studied. Many of the association constants were measured in our laboratory, but most of the data were collected from the literature (*1, 2*). Some of the structures closely resemble natural DNA or RNA base-pairs, others less so, but all compounds contain an approximately linear array of hydrogen-bond donor and acceptor groups. Clefts, macrocycles, and related host-guest systems were not included. The complexes contain up to six primary hydrogen bonds and the free energies of complexation range from ca. −0.8 to 10.9 kcal mol^{-1}. The dataset is believed to be comprehensive and a few representative examples are shown in Figure 1.

Hydrogen Bonding Parameters

Two of the most important predictors of hydrogen bonding strength are proton acidity and number of hydrogen bonds. Amido protons generally form stronger hydrogen bonds than do amino protons because of their higher acidity (*30*). Figure 2 illustrates this effect in a series of acylated diaminopyridines hydrogen bonded to thymine (*31*). Monoacylation of 2,6-diaminopyridine (**7**) leads to a 2.7-fold increase in the association constant. Diacylation increases the association constant by a factor of 5.2 (*32*). In general, base pairs

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005



A·T

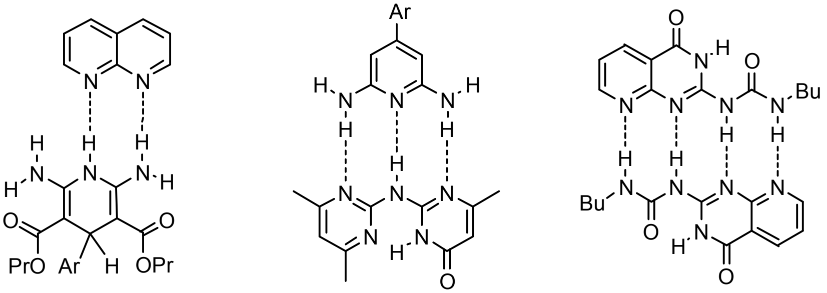
$-\Delta G^{\circ}_{298}$ = 1.74 kcal mol^{-1}
n_{intra} = 0, $n_{NH\cdots N}$ = 1,
$n_{NH\cdots O}$ = 1, $n_{CH\cdots O}$ = 1,
n_{sec} = -3, n_{rot} = 1,
n_{RAHB} = 1

G·C

$-\Delta G^{\circ}_{298}$ = 6.41 kcal mol^{-1}
n_{intra} = 0, $n_{NH\cdots N}$ = 1,
$n_{NH\cdots O}$ = 2, $n_{CH\cdots O}$ = 0,
n_{sec} = 0, n_{rot} = 2,
n_{RAHB} = 3

1·1

$-\Delta G^{\circ}_{298}$ = 2.27 kcal mol^{-1}
n_{intra} = 0, $n_{NH\cdots N}$ = 0,
$n_{NH\cdots O}$ = 2, $n_{CH\cdots O}$ = 0,
n_{sec} = -2, n_{rot} = 0,
n_{RAHB} = 2



2·3

$-\Delta G^{\circ}_{298}$ = 4.7 kcal mol^{-1}
n_{intra} = 0, $n_{NH\cdots N}$ = 2,
$n_{NH\cdots O}$ = 0, $n_{CH\cdots O}$ = 0,
n_{sec} = 3, n_{rot} = 0,
n_{RAHB} = 0

4·5

$-\Delta G^{\circ}_{298}$ = 3.22 kcal mol^{-1}
n_{intra} = 0, $n_{NH\cdots N}$ = 3,
$n_{NH\cdots O}$ = 0, $n_{CH\cdots O}$ = 0,
n_{sec} = -4, n_{rot} = 2,
n_{RAHB} = 3

6·6

$-\Delta G^{\circ}_{298}$ = 10.8 kcal mol^{-1}
n_{intra} = 0, $n_{NH\cdots N}$ = 3.1,
$n_{NH\cdots O}$ = 0.9, $n_{CH\cdots O}$ = 0,
n_{sec} = 2, n_{rot} = 2,
n_{RAHB} = 2

Figure 1. Hydrogen-bonded structures of A·T, G·C, and some synthetic base pairs. Reproduced from reference 1. Copyright 2007 American Chemical Society.

containing a donor site conjugated with a carbonyl group had higher association constants than those without. Within a given series, acylation supplied ca. 0.1 to 1 kcal mol^{-1} stabilization, however, a parameter for this acylation effect was difficult to fit using multiple regression because of the overall variability in complexation strengths. Better fits were obtained when association constants for complexes such as those in Figure 2 were averaged. This has the effect of offsetting differences in acidity of hydrogen bond donors and reducing variability in association constants due to experimental error. In some cases identical hydrogen-bonded complexes measured in different laboratories have association constants differing by up to an order of magnitude. This is presumably due to slightly different experimental conditions from lab to lab.

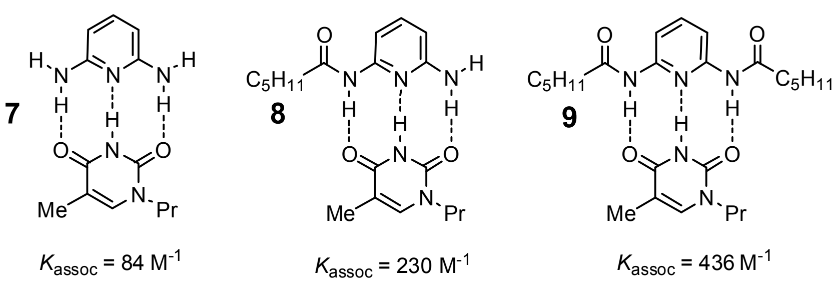


Figure 2. Acylation of diaminopyridine (7) increases K_{assoc} with thymine.

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

Primary and Secondary Interactions

In a preliminary regression analysis that followed Jorgensen's method (*20*), better fits were obtained by using separate variables for the two types of primary hydrogen bonds (NH···O and NH···N). Separating primary O···HN and N···HN hydrogen bonds significantly improved adjusted R^2 values, not only because carbonyl oxygen is a better acceptor than nitrogen, but also because species containing carbonyls generally have more acidic protons.

Repulsive secondary electrostatic interactions occur when adjacent donor or acceptor sites repel each other, leading to destabilization of the complex. Attractive secondary interactions are, essentially, long, non-linear hydrogen bonds that stabilize the complex. An alternating arrangement of hydrogen bonds would therefore lead to a weaker complex. In our analysis, separating the attractive and repulsive secondary interactions had a negligible effect on adjusted R^2. Further separating the components of secondary interactions (O···HN, N···HN, O···O, N···N, H···H) also did not significantly improve adjusted R^2 values. Use of a single variable (n_{sec}) for all the secondary electrostatic components worked as well as the three-component method (repulsive H···H, repulsive N/O···N/O, attractive H···N/O). Thus the single-variable method was chosen for the secondary interactions in the full analysis.

Multiple Regression Analysis

Of the regression analyses performed using all 256 experimental $-\Delta G^{\circ}_{298}$ values, four analyses gave $R^2 > 0.9$, each minimally containing n_{intra}, $n_{\text{NH}\cdots\text{O}}$, $n_{\text{NH}\cdots\text{N}}$, $n_{\text{CH}\cdots\text{O}}$, and n_{sec}. These terms are summarized in Table I. The n_{intra} term is used when an intramolecular hydrogen bond must be broken in one of the components prior to complexation (e.g., pyridylureas) and the term $n_{\text{CH}\cdots\text{O}}$ indicates the number of CH···O contacts. Better fits were obtained when the terms n_{inter} and n_{intra} were separate, rather than combining them as a single term (i.e., n_{inter} minus n_{intra}). Equation 2 fits the data with an adjusted $R^2 = 0.905$. The P values for the intercept and variables were all $< 10^{-6}$, indicating a high degree of statistical significance.

Table I. Summary of Variables in Multivariate Regression Analysis.

Symbol	*Variable*
n_{intra}	Intramolecular H-bonds broken before complexation
$n_{NH\cdots O}$	Intermolecular H-bonds formed between NH and C=O groups
$n_{NH\cdots N}$	Intermolecular H-bonds formed between NH and −N= groups
$n_{CH\cdots O}$	Intermolecular interactions between CH and C=O groups
n_{sec}	Net secondary electrostatic interactions (attractive − repulsive)
n_{RAHB}	Resonance assisted H-bonds (cooperative H-bonds)
n_{rot}	Number of unconstrained single bonds frozen in complex

$$\Delta G^{o}_{298} = 5.6 + 3.2n_{\text{intra}} - 3.5n_{\text{NH}\cdots\text{N}} - 4.1n_{\text{NH}\cdots\text{O}} - 0.7n_{\text{sec}} - 2.2n_{\text{CH}\cdots\text{O}} \qquad (2)$$

When two bases pairs there is a loss of translational and rotation entropy. It is likely that the constant in eq 2, which was necessary for good fits, both accounts for this constant energy loss as well as collecting statistical errors (*29*). Its magnitude (5.6 kcal mol^{-1}) is lower than the value of 7-11 kcal mol^{-1} generally agreed upon as representing the loss in energy when two components are held rigidly fixed (*33*, *34*), but residual motion in complexes will likely lower the value (*35*, *36*). Furthermore, related constants determined in regression analyses of ligand/drug-receptor and enzyme-inhibitor complexes (*37-40*) have similar values. To illustrate the usefulness of the components in eq 2, the best fit of the data using just the total number of primary and secondary hydrogen bonds gave an adjusted $R^2 = 0.554$. Inclusion of an intercept (2.8 kcal mol^{-1}) and excluding those complexes that contain intramolecular hydrogen bonds gave an adjusted R^2 of only 0.750.

Resonance-Assisted Hydrogen Bonds

Resonance-assisted hydrogen bonding (RAHB) occurs when electronic polarization strengthens hydrogen bonds in adjacent sites (*41*). In contrast to the Jorgensen model, RAHB theory implies that an alternating arrangement of donor and acceptor sites enhances complex stability. Addition of the term n_{RAHB} minimally increased the adjusted R^2, as did inclusion of n_{rot} alone or together with n_{RAHB}. Previous studies of RAHBs suggest that it is the σ-skeleton of an unsaturated system that is responsible for the increased stability of intramolecular hydrogen bonds (*42*, *43*). This is consistent with our finding that a term for RAHB is not necessary.

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

Classification of the Base-Pairs

A potential complication in this analysis is the possible tautomerism of the bases studied. In each case we have made the assumption that the tautomeric form is as shown. Additionally, when nondegenerate complexes containing the same hydrogen bonding motif are possible, the K_{assoc} values were statistically corrected. In looking at the data collected, one sees that structurally similar complexes can exhibit significant differences in their K_{assoc} values as a result of differences in the pKa values of the donor and acceptor sites (substituent effects) and because of variability of the actual measurements in different laboratories. To better average out some of these factors the 256 complexes were grouped into 86 structural classes. This approach also helps avoid the over-weighting complex types with multiple entries. Shown in Figure 3a is a plot of $-\Delta G^{o}_{298}$(expt) vs. $-\Delta G^{o}_{298}$(calcd) using eq 2. The grouping of complexes gives, within experimental error, the same equation, but an improved adjusted R^2 value (Figure 3b).

The A·T/U and G·C experimental $-\Delta G^{o}_{298}$ values fit the predicted values from eq 2 within experimental error (standard error 0.65 kcal mol^{-1}) (Table II). Although G·C pairs primarily via Watson-Crick geometry, the A·T/U base pair can exist in four different geometries of approximately equal proportion. For this reason the K_{assoc} for A·T/U was corrected by a factor of four, assuming equal portions of both normal and reversed Watson-Crick and Hoogsteen complexes (Figure 4). Is it not known precisely what the relative amounts of the two forms are in solution although early work on the A·U pair in chloroform-d indicated ca. 70:30 ratio of Hoogsteen to Watson-Crick forms (*44*). More recent studies in $CDClF_2$/CDF_3 suggest stronger bonding in the Watson-Crick arrangement (*45*). Neither study made a distinction between normal and reversed forms.

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

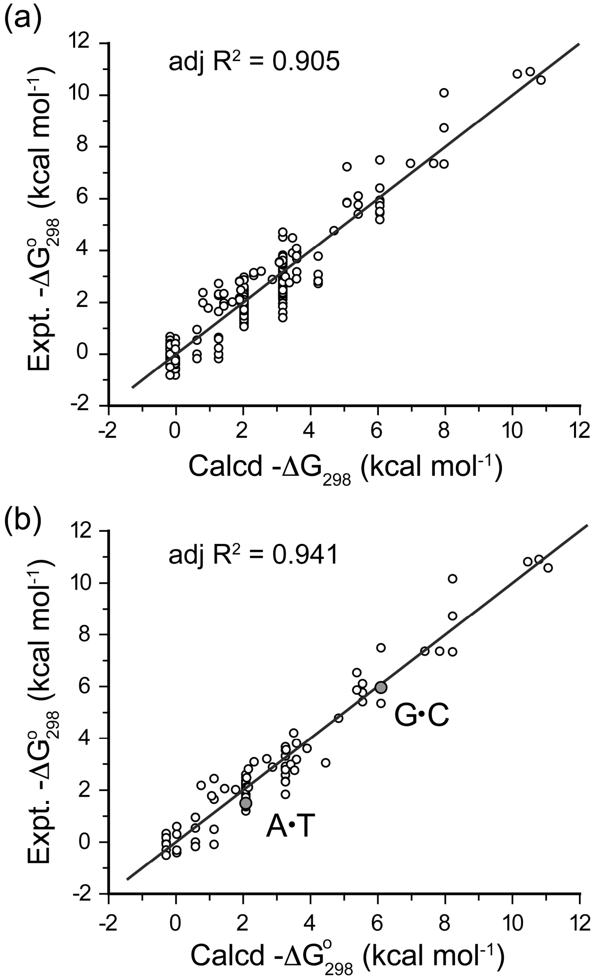
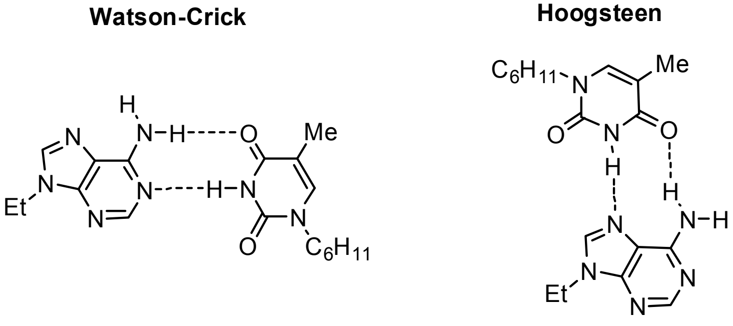


Figure 3. Experimental vs. calculated. $-\Delta G^o_{298}$ using eq 2; (a) 256 separate and (b) 86 grouped complexes. Reproduced from reference 1. Copyright 2007 American Chemical Society

Table II. Comparison of the Experimental and Predicted Values of $-\Delta G^o_{298}$ for the A·T/U and G·C Base Pairs (kcal mol^{-1}).

	Predicted	*Expt. (average)*	*Expt. (range)*
A·U	2.2	1.83	1.666 – 1.962
A·T	2.2	1.63	1.085 – 2.061
G·C	6.1	5.95	5.454 – 6.407

SOURCE: Reproduced from reference 1. Copyright 2007 American Chemical Society.


Watson-Crick

Hoogsteen

Figure 4. Watson-Crick and Hoogsteen pairing geometries for A·T.

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

Ab initio calculations at the MP2/6-31+G(d,p) level indicated that Hoogsteen geometry is preferred over Watson-Crick geometry in A·T by ca. 1 kcal mol^{-1}. Factors that might contribute to the preference for Hoogsteen geometry are a shorter CH···O contact, a favorable alignment of dipoles, and greater distances between secondary repulsive sites. The calculations also indicated that the "reverse" structures are only slightly higher in energy than their respective Watson-Crick and Hoogsteen structures.

CH···O Contacts

A handful of complexes showed experimental stabilities in excess of that predicted by primary hydrogen bonding and secondary interactions alone. In each case a carbonyl group flanks a C-H group. Including an additional parameter in the phenomenological analysis for the putative CH···O interaction led to eq 2 and the excellent fits to the data seen in Figure 3. The role of CH···O hydrogen bonding in base pairs and other complexes has been considered before (*46-53*). Computational studies have examined the contribution of the CH···O interaction to the overall A·T base-pair stability, with values ranging from negligible (*54, 55*) to ca. 6% of the total inter-pair bond energy (*56*). It is well established that CH···O hydrogen bonds exist and they have been discussed in the context of DNA base pairing (*vide supra*). However, our analysis requires an unusually high value ($\geq$ 2.2 kcal mol^{-1}), which is minimally 60% the strength of an average primary hydrogen bond in the model.

The CH···O contact was also investigated in model complexes of adenine with ketene and isocyanic acid. The *ab initio* calculations support the result of the phenomenological approach that the A·T base pair does have enhanced stability relative to hydrogen-bonded complexes with just N–H···N and N–H···O hydrogen bonds. The *ab initio* results for a double-mutant cycle and for the adenine·ketene and adenine·HNCO models certainly suggest that the CH···O interaction plays an important role (though it is a much smaller fraction of the *ab initio* ΔE binding energy of A·T than of the experimental ΔG values), but suggest that it is a group interaction, not a simple CH···O hydrogen bond. This latter conclusion comes from examining the ketene group in a syn and anti arrangement, both of which contain identical CH···O interactions, but lead to quite different stabilizations. More details can be found in reference 1.

Inosine Dimer

The complex that showed the largest deviation in the regression analyses was the inosine dimer (Figure 5), which deviated by more than -2.5 kcal mol^{-1}. Inosine may dimerize via two NH···O hydrogen bonds, however, secondary electrostatic interactions are avoided if the dimer forms via a single NH···N hydrogen bond. In addition, two CH···O contacts are formed (although one is longer than the other by necessity) and the inosine dimer fits the model within 0.7 kcal mol^{-1}. Crystallographic data (*57-59*) indicate single hydrogen bond geometry, although solid-state structures are not always indicative of solution structure. A dilution study was performed to test experimentally whether inosine pairs via two NH···O hydrogen bonds or via one NH···N hydrogen bond and two CH···O contacts. Upon diluting a chloroform-d solution from 75 mM to 0.08 mM, the chemical shift of H2 changed by -0.330 ppm and H8 changed by -0.049 ppm. The larger upfield shift of H2 relative to H8 may indicate a stronger H2···O interaction or a contribution from both hydrogen-bonded species in Figure 5.

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

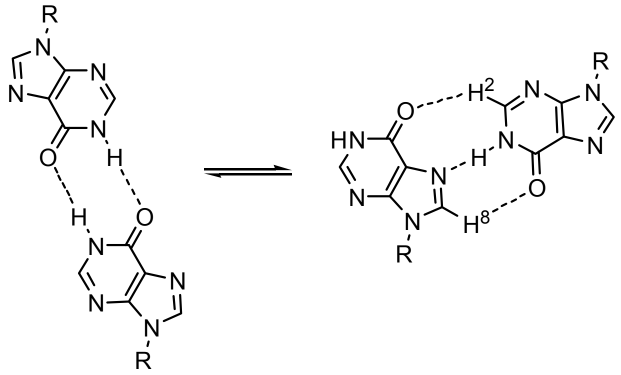


Figure 5. The inosine dimer pairs via two NH···O hydrogen bonds (lefthand structure) or via a single NH···N hydrogen bond and two CH···O contacts (righthand structure).

Conclusions

A goal of this work was to identify any special stability in the natural base pairs. Although the G·C base pair has a very high association constant, it displays no special stability relative to other complexes containing N−H···N and N−H···O hydrogen bonds. The A·T base pair, on the other hand, requires an additional parameter, a CH···O contact, to fit our model. Our combined experimental and computational results indicate that this putative CH···O contact is worth about ca. 2−2.5 kcal mol^{-1}, though it's unlikely this extra stability arises from a true CH···O hydrogen bond. These findings demonstrate the plausibility that the A·T base pair was propagated in biological systems because of its unique stability.

It is important to note that the excellent fit generated from eq 2 does not confirm the physical significance of the individual components, rather it establishes a correlation between assignable increments and complexation free energy. Other factors may play a role, but are not easily assigned variables. In any case, this model highlights primary hydrogen bonding and secondary interactions as well as intramolecular hydrogen bonding in pyridylurea-type complexes as significant indicators of hydrogen bonding strength. Equation 2 has predictive value and suggests CH···O contacts as a useful design consideration for synthetic bases.

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

References

1. This Chapter was adapted from: Quinn, J. R.; Zimmerman, S. C.; Del Bene, J.; Shavitt, I. *J. Am. Chem. Soc.* **2007,** *129,* 934−941.
2. Quinn, J. R. Ph.D. thesis, University of Illinois, Urbana, IL, 2006.
3. Woese, C. *The Genetic Code. The Molecular Basis for Genetic Expression;* Harper-Row: New York, 1967; pp 179−195.
4. Crick, F. H. C. *J. Mol. Biol.* **1968,** *38,* 267−379.
5. Orgel, L. E. *J. Mol. Biol.* **1968,** *38,* 381−393.
6. Gilbert, W. *Nature (London)* **1986,** *319,* 618.
7. Reader, J. S.; Joyce, G. F. *Nature (London)* **2002,** *402,* 841−844.
8. Shapiro, R. *Orig. Life Evol. Biosphere* **1995,** *25,* 83−98.
9. Oró, J. *Biochem. Biophys. Res. Commun.* **1960,** *2,* 407−415.
10. Sanchez, R.; Ferris, J.; Orgel, L. E. *Science* **1966,** *153,* 72−73.
11. Saladino, R.; Crestini, C.; Costanzo, G.; Negri, R.; Di Mauro, E. *Bioorg. Med. Chem.* **2001,** *9,* 1249−1253.
12. Zubay, G. *Orig. Life Evol. Biosphere* **2001,** *31,* 87−102.
13. Eschenmoser, A. *Tetrahedron* **2007,** *63,* 12821−12844.
14. Mittapalli, G. K.; Osornio, Y. M.; Guerrero, M. A.; Reddy, K. R.; Krishnamurthy, R.; Eschenmoser, A. *Angew. Chem. Int. Ed.* **2007,** *46,* 2478−2484.
15. Saenger, W. *Principles of Nucleic Acid Structure; Springer-Verlag;* New York, 1984; chapter 6.

16. Kyogoku, Y.; Lord, R. C.; Rich, A. *Proc. Natl. Acad. Sci., U.S.A.* **1967,** *57,* 250−257.
17. Kyogoku, Y.; Lord, R. C.; Rich, A. *Biochim. Biophys. Acta* **1969,** *179,* 10−17.
18. Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990,** *112,* 2008–2010.
19. Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991,** *113,* 2810–2819.
20. Jorgensen, W. L. *Chemtracts Org. Chem.* **1991,** *4,* 91–119.
21. Jeong, K. S.; Tjivikua, T.; Muehldorf, A.; Deslongchamps, G.; Famulok, M.; Rebek, J. Jr. *J. Am. Chem. Soc.* **1991,** *113,* 201−209.
22. Leonard, G. A.; McAuley-Hecht, K.; Brown, T.; Hunter, W. N. *Acta Crystallogr., Sect. D* **1995,** *51,* 136−139.
23. Murray, T. J.; Zimmerman, S. C. *J. Am. Chem. Soc.* **1992,** *114,* 4010−4011.
24. Sartorius, J.; Schneider, H.-J. *Chem. Eur. J.* **1996,** *2,* 1446−1452.
25. Beijer, F. H.; Sijbesma, R. P.; Kooijman, H.; Spek, A. L.; Meijer, E. W. *J. Am. Chem. Soc.* **1998,** *120,* 6761−6769.
26. Lüning, U.; Kühl, C. *Tetrahedron Lett.* **1998,** *39,* 5735−5738.
27. Zimmerman, S. C.; Corbin, P. S. *Struct. Bonding* **2000,** *96,* 63−94.
28. Alvarez-Rua, C.; García-Granda, S.; Goswami, S.; Mukherjee, R.; Dey, S.; Claramunt, R. M.; Santa María, M. D.; Rozas, I.; Jagerovic, N.; Alkorta, I.; Elguero, J. *New J. Chem.* **2004,** *28,* 700−707.
29. Hair, J. F.; Anderson, R. E.; Tatham, R. L.; Black, W. C. *Multivariate Data Analysis;* Prentice Hall: New York, 1998.
30. Beijer, F. H.; Sijbesma, R. P.; Vekemans, J. A. J. M.; Meijer, E. W.; Kooijman, H.; Spek, A. L. *J. Org. Chem.* **1996,** *61,* 6371−6380.
31. Yu, L.; Schneider, H.-J. *Eur. J. Org. Chem*. **1999,** 1619−1625.
32. Söntjens, S. H. M.; Meijer, J. T.; Koojiman, H.; Spek, A. L.; van Genderen, M. H. P.; Sijbesma, R. P.; Meijer, E. W. *Org. Lett.* **2001,** *3,* 3887−3889.
33. Page, M. I; Jencks, W. P. *Proc. Natl. Acad. Sci., U.S.A.* **1971,** *68,* 1678−1683.
34. Ajay; Murcko, M. A. *J. Med. Chem.* **1995,** *38,* 4953−4967.
35. Searle, M. S.; Williams, D. H. *J. Am. Chem. Soc.* **1992,** *114,* 10690−10697.
36. Searle, M. S.; Williams, D. H.; Gerhard, U. *J. Am. Chem. Soc.* **1992,** *114,* 10697−10704.
37. Horton, N.; Lewis, M. *Protein Sci.* **1992,** *1,* 169−181.
38. Krystek, S.; Stouch, T.; Novotny, J. *J. Mol. Biol.* **1993,** *234,* 661−679.
39. Böhm, H.-J. *J. Comput.-Aided Molec. Design* **1994,** *8,* 243−256.
40. Head, R. D.; Smythe, M. L; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. *J. Am. Chem. Soc.* **1996,** *118,* 3959−3969.
41. Gilli, G.; Bellucci, F.; Ferretti, V.; Bertolasi, V. *J. Am. Chem. Soc.* **1989,** *111,* 1023−1028.
42. Alkorta, I.; Elguero, J.; Mó, O.; Yañez, M.; Del Bene, J. E. *Mol. Phys.* **2004,** *102,* 2563−2574.
43. Alkorta, I.; Elguero, J.; Mó, O.; Yañez, M.; Del Bene, J. E. *Chem. Phys. Lett.* **2005,** *411,* 411−415.
44. Iwahashi, H.; Sugeta, H.; Kyogoku, Y. *Biochemistry* **1982,** *21,* 631−638.

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

45. Dunger, A.; Limbach, H.-H.; Weisz, K. *J. Am. Chem. Soc.* **2000,** *122,* 10109−10114.
46. Gu, Y.; Kar, T.; Scheiner, S. *J. Am. Chem. Soc.* **1999,** *121,* 9411–9422.
47. Scheiner, S. *Adv. Mol. Struc. Res.* **2000,** *6,* 159–207.
48. Scheiner, S.; Gu, Y.; Kar, T. *J. Mol. Struc. (Theochem)* **2000,** *500,* 441–452.
49. Gu, Y.; Kar, T.; Scheiner, S. *J. Mol. Struc.* **2000,** *552,* 17–31.
50. Scheiner, S.; Kar, T.; Gu, Y. *J. Biol. Chem.* **2001,** *276,* 9832–9837.
51. Scheiner, S.; Grabowski, S. T.; Kar, T. *J. Phys Chem. A* **2001,** *105,* 10607–10612.
52. Brandl, M.; Meyer M.; Suhnel, J. *J. Biomol. Struct. Dyn*. **2001,** *18,* 545–555.
53. Hartmann, M.; Wetmore, S. D.; Radom, L. *J. Phys. Chem. A* **2001,** *105,* 4470–4479.
54. Asensio, A.; Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003,** *107,* 6441–6443.
55. Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem. Eur. J.* **1999,** *5,* 3581–3594.
56. Starikov, E. B.; Steiner, T. *Acta Crystallogr., Sect. D* **1997,** *53,* 345–347.
57. Mande, S. S.; Shamala, N.; Seshadri, T. P.; Viswamitra, M. A. *Nucleosides Nucleotides* **1992,** *11,* 1103–1114.
58. Hakoshima, T.; Omori, H.; Tomita, K.; Miki, H.; Ikehara, M. *Nucleic Acids Res.* **1981,** *9,* 711–729.
59. Kolappan, S.; Seshadri, T. P. *Acta Crystallogr., Sect. C* **1999,** *55,* 986–987.

Downloaded by PENNSYLVANIA STATE UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch005

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

Chapter 6

Self-Assembly and the Origin of the First RNA-Like Polymers

Heather D. Bean[1,2], David G. Lynn[1,3], and Nicholas V. Hud[1,2]

[1]Center for Fundamental and Applied Molecular Evolution, Atlanta, GA
[2]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332
[3]Departments of Chemistry and Biology, Emory University, Atlanta, GA 30322

Many hypotheses for the origin of life rely on the existence of an RNA world, a time in which RNA both stored genetic information and performed catalysis, functions that are performed by DNA and proteins, respectively, in extant biology. However, the *de novo* synthesis of RNA by plausible prebiotic reactions has yet to be demonstrated, leading many researchers to conclude that RNA was preceded by an informational polymer that resembled RNA, but was easier to assemble, i.e., proto-RNA. Still, the synthesis of a proto-RNA is not trivial, requiring the selection of a subset of building blocks out of a diverse prebiotic chemical inventory, and their correct coupling into a polymer. In this chapter we focus on the difficulty of selection and coupling in the *de novo* synthesis of RNA-like polymers and provide support for the utility of molecular midwives and reversible bonding along a template in overcoming these challenges.

The RNA World – A Bridge to the Origin of Life?

All known life forms on Earth utilize an intricate and interdependent system of biopolymer synthesis and cellular function: genetic information is stored in DNA, which is transcribed into RNA that is then decoded by the ribosome and translated into protein. The universality of this system, the "Central Dogma" of

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

molecular biology, has led to the hypothesis that all extant biology is descended from a single organism, or pool of interrelated organisms, that utilized DNA for information storage and proteins as reaction catalysts. However, with respect to the origin of life, the hypothesis of a DNA/protein-based common ancestor presents a classic chicken-or-the-egg paradox, "Which came first, DNA or proteins?" DNA codes for protein synthesis, yet proteins are required for nucleic acid synthesis and replication. Proposed in the late 1960s, the RNA world hypothesis provided a solution – the first macromolecule was neither DNA nor protein, but rather RNA (*1-3*). As the intermediary molecule between information storage and protein synthesis, RNA is a possible candidate to be the progenitor of both DNA and protein. Additionally, RNA is capable of storing genetic information in its base sequence, and catalyzing chemical reactions. In fact, examples of the catalytic role of RNA in extant biology are now known to range from self-splicing mRNA molecules to the ribosomal RNA nucleotides that directly catalyze the coupling of amino acids in a growing peptide chain (*4*).

The synthesis of RNA in extant biology, however, still relies upon the participation of proteins. The protein-free *de novo* synthesis of RNA in a prebiotic reaction has yet to be demonstrated after several decades of effort (*5*). Many researchers have therefore concluded that RNA was not the first informational polymer of life. Rather, RNA was preceded by an RNA-*like* polymer, or several generations of polymers, termed proto-RNAs, that were structurally and functionally similar to RNA, but easier to assemble. Proto-RNA could have been comprised of different bases, sugars, and linking molecules that were assembled through more thermodynamically and kinetically accessible pathways. Without the constraints of the current four RNA bases, ribose, and phosphate for the construction of an informational polymer, the possible composition of proto-RNAs seems limitless. However, the existence of putative monomer units in the prebiotic chemical inventory for the assembly of proto-RNA would have been dictated by astro- and geochemical processes, paring down the set of molecules from which nature could select.

The origin of life community is compiling an ever-growing list of compounds that are believed to have been available for prebiotic chemical reactions. Astronomers report the compositions of neighboring terrestrial bodies' atmospheres and surfaces, meteorites, comets, and the seemingly empty interstellar medium (*6-8*), which contain a diversity of molecules ranging from molecular hydrogen to ethanol to cyanodecapentayne ($HC_{11}N$) (*9*). Geologists and atmospheric chemists hypothesize about the composition of the early Earth's atmosphere, oceans, and land based on mathematical models and minerals comprising the few ancient rocks that have survived the metamorphosis of the planet (*10,11*). Laboratory chemists use these molecules and proposed environmental conditions to guide model prebiotic reactions to further expand our understanding of the prebiotic chemical inventory (*12-19*). Although plausible prebiotic syntheses for many biological building blocks have been demonstrated (e.g., ribose, several amino acids, and the RNA bases), a prebiotic methodology for coupling the RNA monomers into an oligonucleotide has not been demonstrated. Thus, there is currently a gap in our conceptual evolutionary timeline of life from the small molecule world (i.e., the prebiotic chemical inventory) to the appearance of the first proto-RNA molecule.

The Missing Link

The gap in the prebiotic timeline can be attributed to two distinct challenges, selection and coupling. From the point of view of selection alone, the synthesis of RNA from the prebiotic chemical inventory is a formidable challenge. From the vast prebiotic chemical inventory that likely contained innumerable molecules resembling the building blocks of RNA, how were the Watson-Crick bases and ribose selected? (Figure 1) Even if a process existed by which the monomer units could be selected from a complex chemical mixture, how did these building blocks become coupled specifically into β-furanosyl nucleosides, phosphorylated, and incorporated into an RNA polymer in a 3′,5′ orientation so as to assemble extant RNA? (Figure 1) Although evoking the existence of a proto-RNA world helps relax some of the restrictions for selection and coupling, these two steps still have a profound influence on the synthesis of the original informational polymers. In the following sections we will elaborate on the challenges of selection and coupling as they relate to *de novo* RNA synthesis in the prebiotic environment.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

Nucleoside Formation

As discussed above, researchers have defined a varied and ever-growing prebiotic chemical inventory, from which adenine, guanine, cytosine, uracil and D-ribose would have had to have been selected for RNA synthesis. On what criteria were these selections made? One possible answer is that the selection was primarily based on high concentrations of RNA precursors relative to similar molecules in the prebiotic environment. However, several lines of evidence suggest that this mode of selection was not applicable to prebiotic RNA synthesis, as demonstrated by the case of D-ribose. The formose reaction famously produces an incredible range of sugar products from glyceraldehyde to hexoses, branched and straight-chain, with every possible combination of stereoisomer and handedness (D and L) observed in the mixture (*20*). It could be argued that the lack of selectivity in ribose production could have been overcome if ribose was significantly more stable than the other sugars produced by the formose or other reaction, thereby concentrating ribose in the mixture over time. However, the Miller laboratory demonstrated that the half-life of ribose in aqueous solution is within an order of magnitude of the half-lives of other aldopentoses and aldohexoses (*21*). Therefore, it could not have been the concentration of ribose in solution, either through synthesis or stability, that selected it for inclusion in RNA molecules.

The argument of selective synthesis in regards to the incorporation of β-furanosyl nucleosides as the sole building monomer unit in RNAs is also disproved. When coupling bases and ribose together in a simple heating-drying reaction (*22,23*), four nucleoside anomers are possible, β-furanosyl, β-pyranosyl, α-furanosyl, and α-pyranosyl ribonucleosides (Figure 2), and the β-furanosyl nucleoside is produced without selectivity (*22-26*). This phenomenon is not restricted to glycosides of ribose, but is common for any sugar that adopts multiple conformations in solution (25,*27-29*).

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

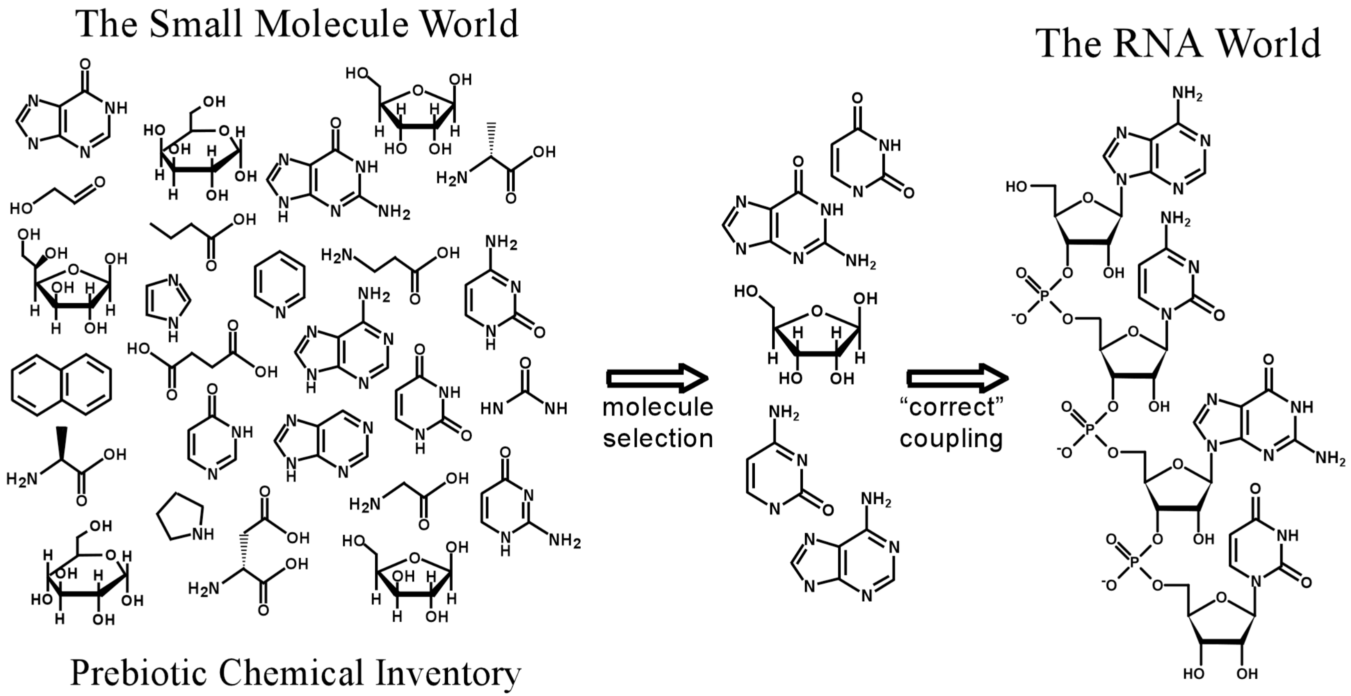

The Small Molecule World
Prebiotic Chemical Inventory
molecule selection
"correct" coupling
The RNA World

Figure 1. From the small molecule world to the RNA world. The two primary challenges for assembling RNA (or proto-RNA) from the prebiotic chemical inventory are the selection of the correct RNA monomers (adenine, uracil, guanine, cytosine, and ribose), and coupling of the monomers in the correct way into nucleosides, nucleotides, and oligonucleotides.

It has also been demonstrated that the distribution of nucleoside anomers produced in acidic aqueous solutions is kinetically controlled on short time scales, but will reach an equilibrium distribution that minimizes steric interactions between the base and sugar, which for aldopentoses results in the predominance of the α- and β-pyranosyl forms (*25,27-29*). Therefore, it would not have been the selective production of β-furanosyl nucleosides that influenced their incorporation into prebiotic nucleic acids.

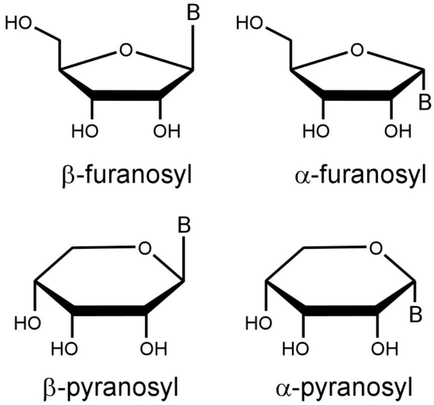


Figure 2. Four anomers of ribosyl nucleosides. "B" represents the nucleoside base, which could be uracil, cytosine, adenine, or guanine for RNA.

Eschenmoser and co-workers have sought to understand if there is a structural or chemical advantage that β-furanosyl ribose imposes on RNA that selected it from the anomers of ribose and other sugars for the construction of the backbone. Their data show that there are several sugars (tetroses, pentoses, hexoses) and alternative conformations of ribose (e.g., β-pyranosyl) that are capable of forming double-stranded structures when incorporated into nucleic acid polymers; some of the modified duplexes are even more stable than the RNA duplex (*30-34*). Taken together, these data support the notion that the selective formation of a *ribo*nucleoside out of the prebiotic chemical inventory is not likely, and that proto-RNA backbones may have been constructed from a sugar (or a mixture of sugars) other than ribose.

Coupling and the Nucleoside Problem

In 1972, Fuller and Orgel reported the first prebiotic synthesis of ribonucleosides, coupling the preformed bases with ribose in a drying-heating reaction (*22*). This method of synthesis was successful for the formation of adenosine and inosine from their concomitant purine bases, adenine and hypoxanthine, but did not work for the synthesis of uridine or cytidine (*22,23*). Two years earlier Sanchez and Orgel had demonstrated the stepwise synthesis of cytidine monophosphate, beginning with 5′-phosphorylated ribose, followed by multiple steps, including chemical addition, UV anomerization, and hydrolysis (*35*). Despite three decades of experimentation by Orgel and others to find a

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

more plausible prebiotic route to uridine and cytidine, no simpler methods have arisen (*36-38*). The lack of pyrimidine nucleoside synthesis, known as the nucleoside problem, has even lead some scientists to hypothesize that pyrimidines were absent from the earliest nucleic acid polymers (*2,3,39-41*).

Our lab has recently proposed that the barrier to uridine and cytidine synthesis lies in the nucleophilicity of the base (*26*). The absence of an in-plane lone pair on N1 dramatically slows nucleophilic attack on C1′ of the sugar (Figure 3). Deprotonation at N1 of uracil or cytosine (Figure 3a) would require an alkaline pH, which is incompatible with the acid-catalyzed step of glycosidic bond formation. Indeed, N9 of adenine and hypoxanthine are sufficiently nucleophilic to undergo glycosylation via our proposed mechanism in the reported syntheses of Orgel and co-workers (*22*). The successful synthesis of purine nucleosides suggests that there may be pyrimidine bases other than uracil and cytosine that are better suited for non-enzymatic pyrimidine nucleoside formation, given appropriate substitutions on the base. This hypothesis was confirmed by coupling 2-pyrimidinone to ribose under Orgel's reaction conditions and with good yield (Figure 3b) (*26*). As 2-pyrimidinone is a pyrimidine base, this result demonstrates that pyrimidine nucleosides are prebiotically plausible, but perhaps uridine and cytidine were not available for inclusion in the earliest proto-RNA molecules.

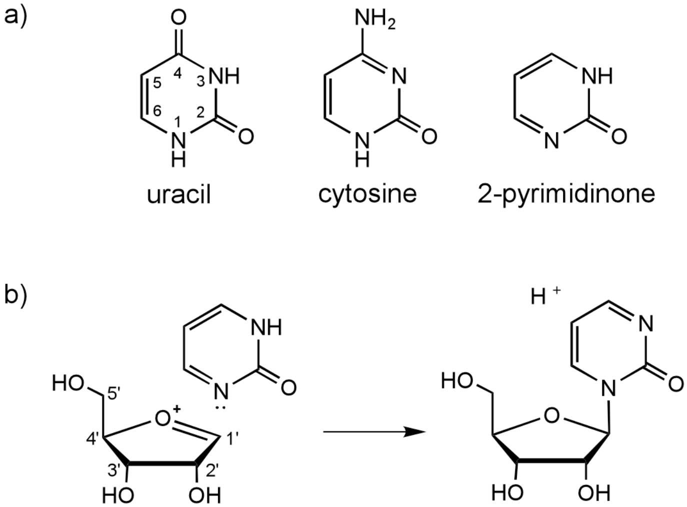


Figure 3. Proposed mechanism of pyrimidine nucleoside formation. a) Structures of the pyrimidine bases uracil, cytosine, and 2-pyrimidinone. Uracil and cytosine are protonated at N1 and therefore do not possess an in-plane lone pair of electrons, whereas 2-pyrimidinone does. b) The lone pair on 2-pyrimidinone nucleophilically attacks C1′ of the sugar oxonium intermediate to form the glycosidic bond.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

The Paradox of Base Pairing

Selection of the nucleic acid bases (A, G, C, U) for prebiotic RNA synthesis also poses a barrier. Although the Watson-Crick base pairs are energetically favored within a duplex, individual bases will not form hydrogen bonds in aqueous solution, nor will their nucleosides or nucleotides, but rather the bases will form columnar stacks (*42*). Stacking is thermodynamically favored over base pairing as it reduces the solvent-exposed surface area of the largely hydrophobic bases, and the hydrogen donor and acceptor groups that participate in base pairing can be satisfied through H-bonding with water. Despite the fact that Watson-Crick base pairs are essential to information transfer and storage in all organisms, and therefore were likely selected by biology's earliest ancestors explicitly because of this pairing ability, the free bases do not form base pairs in aqueous solution. This realization elicits the *Paradox of Base Pairing* (*43*) – How would the bases have been selected for inclusion in the first proto-RNA molecules for the purpose of Watson-Crick base pairing if the bases themselves did not form base pairs without having first been linked by a backbone?

Even when incorporated into a polymer, pairing of the four RNA bases is not restricted to the Watson-Crick faces. For example, the guanine-uracil wobble base pair is relatively stable and well represented in ribonucleic acid secondary structures, and non-Watson-Crick (e.g., Hoogsteen) pairing occurs in triplex and quadruplex nucleic acid structures. As with the synthesis of sugars, the prebiotic formation of the bases was most likely not limited to the two purines and two pyrimidines of RNA. Rather, a myriad of similar heterocyclic molecules were potentially available in the prebiotic chemical inventory (*18*,*24*). Many examples exist of modified nucleic acid base pairs that function within a ribonucleic acid duplex, including the modified bases of tRNA, the size-expanded duplexes of Kool and co-workers, and even purine-purine duplexes (*41,44,45-47*). With such a diversity of nucleic acid base-like molecules with multiple pairing conformations available, it is difficult to imagine how the earliest proto-RNAs would have spontaneously incorporated bases that participate in the formation of Watson-Crick base pairs without there having first existed a mechanism that selected the bases that can form such base pairs from the myriad of molecules that do not.

Nucleotide and Oligonucleotide Synthesis

The RNA backbone consists of repeating units of ribose and phosphate linked through phosphodiester bonds. The advantages that phosphate imparts to RNA are numerous, not the least of which are hydrolytic protection, aqueous solubility, maintenance of an extended backbone conformation, and genotypic-phenotypic decoupling, all provided by the negative charge of phosphate at neutral pH (*48,49*). However, the ubiquity of phosphate nucleotides in biology, not just as nucleic acid monomers, but also as the energy currency of the cell (i.e., ATP), derives from the juxtaposition of thermodynamic instability and kinetic stability of phosphoesters and pyrophosphates (*48*). Alternative possibilities can be imagined for nucleic acid linking molecules that also possess

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

an acid group, providing the benefits of negative charges (*48,50*), but it is the thermodynamic and kinetic properties of phosphate that likely drove its evolutionary selection. However, these same attributes may have also excluded phosphates from early proto-RNA molecules.

The hypothesis that phosphate was not the linker molecule in early predecessors of RNA is not new, and is partly driven by the lack of phosphate in available forms on the Earth (*51*). Orthophosphate, the predominant phosphorous-containing moiety in extant biology, is a limiting reagent for living organisms in the environment because of its low solubility. The vast majority of phosphate on the Earth is mineralized, and the hypothesized atmospheric and oceanic conditions of the early Earth would have also resulted in low abundances of available orthophosphate (*51*). Although recent studies have demonstrated the ability of phosphate minerals to phosphorylate nucleosides in formamide (*52*), and new hypotheses are arising for the availability of soluble reduced phosphorous compounds on the early Earth (*53,54*), phosphoester bond formation is still an enthalpically unfavorable process.

The thermodynamic impediment of phosphodiester formation restricts the ability to test hypotheses regarding the driving factors of nucleotide polymerization and replication (e.g., sequence effects, stacking interactions, surface adhesion, etc.). Therefore, many researchers have chosen to use one of a variety of available activating chemistries to drive phosphodiester bond formation, thereby removing the activation step as the limiting factor of oligonucleotide synthesis or replication. However, the use of activating agents in phosphodiester bond formation leads to backbones with mixed 3′,5′ and 2′,5′ linkages, misincorporated (i.e., non-W-C pairing) bases, and truncation products resulting from strand cyclization (*5,55,56*). In the absence of an editing mechanism, such as the enzymatic processes involved in cellular DNA replication, these undesirable products represent kinetic traps. As a result, particularly in the case of cyclized products, the yield of long oligonucleotides is greatly affected. For example, Ferris and coworkers have observed that using 1-methylimidazole activated adenine mononucleotides for oligomerization on mineral surfaces results in a large proportion of cyclic dimer formation (*56*). The problem of cyclization is not restricted to non-templated reactions or to very short polymers (i.e., dinucleotides), but is also observed as a major side product in carbodiimide-facilitated ligation of tiled half-complementary sixmer oligonucleotides (*55*). Beyond being a nuisance in the lab, kinetically-trapped byproducts would have also halted polymerization and replication of nucleotides in the prebiotic environment.

Sequence effects also present a significant barrier to non-enzymatic template directed replication of nucleic acids. As detailed above with the paradox of base pairing, hydrogen bonding of the template strand with the complementary monomer in water is not a sufficient driving force for monomers to associate with the primer of a replicating strand. Rather, monomer association is governed by base stacking. As the pyrimidines have poor stacking interactions compared to those of purines, they are not efficiently incorporated into the product strands of templated reactions. In fact, two or more sequential adenines in an oligonucleotide template strand will halt replication as the self-association constant for thymidine is too low for tandem incorporation into the product

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

strand (*57*). Other problems such as the formation of G-quadruplexes by G-rich template strands, and non-Watson-Crick base pairings (e.g., G-U wobble pairing) further inhibit high-fidelity sequence transfer in template-directed synthesis. These problems would greatly restrict the sequence space for evolving ribozymes and proto-ribozymes as newly mutated sequences that possess new catalytic properties may not be replicable (i.e., heritable).

Non-enzymatic template-directed synthesis is also plagued by the low association constants of monomer nucleotides or short oligonucleotide sequences for the template strand. Mineral surfaces have long been touted as a means of locally concentrating the monomers of biomolecules on the early Earth (*58*), and they have been demonstrated to enhance the polymerization of activated nucleotide monophosphates (*59*). However, no mineral surface has been shown to bind the RNA bases in an orientation that allows base pairing during polymer formation, nor been demonstrated to be selective for the RNA bases, and most would likely adsorb a range of compounds. As the prebiotic chemical inventory is hypothesized to have contained a wide variety nucleic acid base-like heterocycles of various sizes, shapes, and proton donor and acceptor patterns, the non-selective adsorption of these molecules onto a mineral surface, followed by their inclusion into a polymer, would likely result in an oligonucleotide that is unable to base-pair, and therefore unable to replicate. These and other characteristics of mineral-catalyzed polymerization, in particular irreversible product strand binding (*60*), make mineral surfaces of questionable utility as promoters of oligonucleotide synthesis and replication in the prebiotic environment.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

Bridging the Gap

As we have outlined above, bridging the gap between the small molecule world and the RNA world (or proto-RNA world) requires a means of selecting and coupling base pairs into an oligonucleotide. We propose two complementary solutions to the quandaries of selection and coupling for the synthesis of proto-RNAs, molecular midwives and reversible coupling chemistries, which are elaborated below.

Molecular Midwives: Base Selection and Preorganization

We propose that there was a template upon which the nucleic acid base pairs were preorganized prior to their incorporation into oligonucleotides to form proto-RNAs. These templates would have been heterocyclic molecules available in the prebiotic chemical inventory, produced by reactions and starting materials similar to those that made the nucleic bases in the small molecule world. We have used the term ‘molecular midwife’ to describe any such molecule that facilitated the formation of proto-RNA, but was not covalently linked to the polymer (*61*). These midwife molecules would have aided in the ‘birth’ of a proto-RNA, but would have no longer been necessary for RNA synthesis and replication once evolution had produced a superior means for carrying out these reactions (e.g., catalysis via ribozymes or protein enzymes).

We envision that an example molecular midwife would select base-pairing molecules from the mixture of heterocyclic compounds by providing a nanometer scale template upon which the base pairs could stack in solution (Plate 1). The midwife molecules would have been similar in size and shape to the small planar molecules known to intercalate the bases of extant RNA and DNA (*62*). Through the formation of a columnar stack of alternating midwifes and base pairs, the bases would be locally concentrated and oriented in the correct position for joining by the formation of a polymer backbone. As the bases would be spaced at 6.8 Å by the intervening midwife molecules, the backbone that stitched the bases together would resemble the length of the extant RNA backbone, allowing the nascent proto-RNA polymer to form a helical duplex upon removal of the intercalating midwife molecules, and would allow the single strands of proto-RNA the flexibility to form folded structures that are a common characteristic of ribozymes. Additionally, the selection and organization of the bases as base pairs ensures that the proto-RNA polymer can be replicated via the synthesis of a complementary strand, and also ensures the *de novo* synthesis of polymers that form a duplex of homogeneous width, required for duplex stability (*45*). Replication facilitated by molecular midwives could also alleviate some template sequence effects that plague non-enzymatic template directed replication, such as the inability to replicate purine-rich templates due to the low self-association constants of pyrimidine nucleotides (*57*). As the midwife molecules are not covalently linked to the newly synthesized polymers, their dissociation (and subsequent reassociation for a new synthesis/replication cycle) is easily facilitated through changes in solution condition, such as concentration, ionic strength, or temperature (Plate 1).

The association constants of intercalators for DNA or RNA generally range between $10^5 - 10^6$ M^{-1} (*63*). Therefore, if the concentration of intercalator were increased above approximately 10 μM in the presence of nucleic acid bases, the equilibrium amount of the bases that are assembled with intercalators (versus free in solution) would increase. Templating could occur not only for duplexes, but also for triplexes or quadruplexes, given a molecular midwife of the appropriate size and shape (Figure 4). We have demonstrated the power of intercalators to promote duplex- and triplex-based template-directed syntheses in previous reports (*64,43*). In these works, proflavine and coralyne were shown to facilitate up to a 1000-fold enhancement in the coupling of dT_3 and dT_4

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

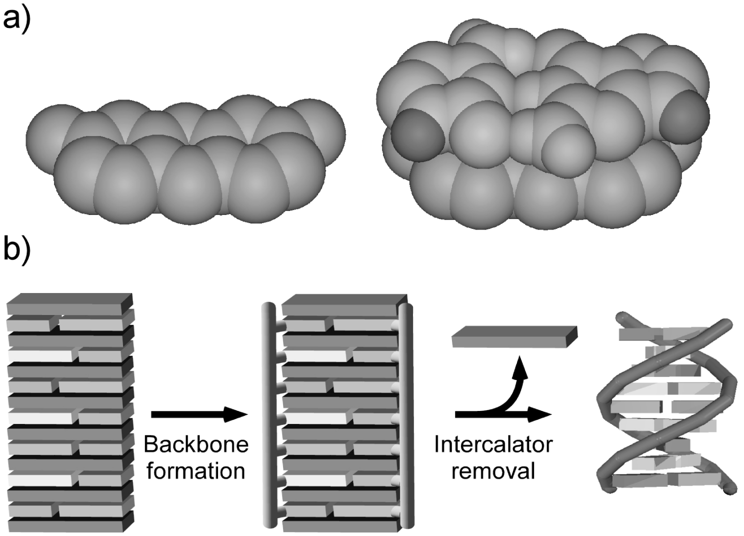


Plate 1.(see color insert 1) Schematic representation of the 'molecular midwife' hypothesis (57). a) A midwife molecule, proflavine (in blue) acts as a template for the formation of a Watson–Crick base pair. H-atoms of the nucleoside bases that will be replaced by C-atoms upon backbone formation are shown in magenta. b) In the case of molecular midwives with a greater association constant for base assembly than for self association, columnar stacks containing alternating midwife molecules and base assemblies spontaneously form under certain conditions. These columnar stacks preorganize the bases such that the introduction of a linkage chemistry leads to the formation of RNA-like polymers with a backbone length of 6.8 Å. Because the midwife molecules are only associated with the resulting polymers through non-covalent interactions, changes in solution conditions can lead to the removal of the intercalating midwives. (Reproduced with permission from reference 41. Copyright 2007 Verlag Helvetica Chimica Acta.)

phosphorothioate substrate strands into a dT_7 product in the presence of an existing template strand (proflavine, Figure 5a) or an existing hairpin template (coralyne, Figure 5b). Assembly of the substrate strands on the template oligonucleotides was also shown to be shape-specific, as proflavine, with the shape of a Watson-Crick base pair, is able to promote template-directed synthesis in a duplex-based system but not in a triplex-based system, while the opposite is true for coralyne, which has the shape of a pyrimidine triplet (Figure 4) (*43*). Although the coupling chemistry used in these ligation experiments is not considered to be prebiotically plausible, the results demonstrate that the assembly of an otherwise unstable template and substrate complex is promoted by intercalation, and that backbone coupling is possible in an intercalation-stabilized assembly.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

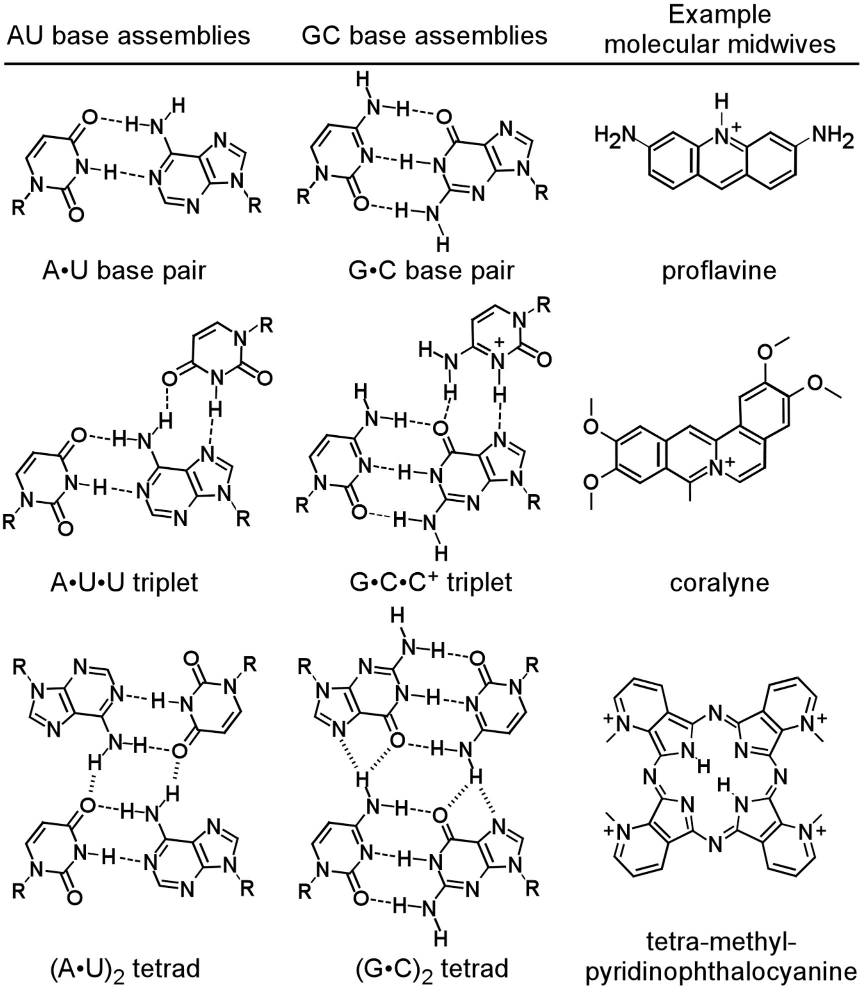

AU base assemblies
GC base assemblies
Example molecular midwives
A•U base pair
G•C base pair
proflavine
A•U•U triplet
G•C•C+ triplet
coralyne
(A•U)2 tetrad
(G•C)2 tetrad
tetra-methyl-pyridinophthalocyanine

Figure 4. Nucleoside base assemblies and heterocyclic molecules with similar shapes. Proflavine and coralyne have already been shown to act as midwifes in ligation assays that include duplex or triplex formation between the template and substrate strands, respectively. The phthalocyanine analogue with a similar shape to the base tetrads is shown for illustrative purposes, as a molecular midwife has not yet been demonstrated for a tetrad system.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

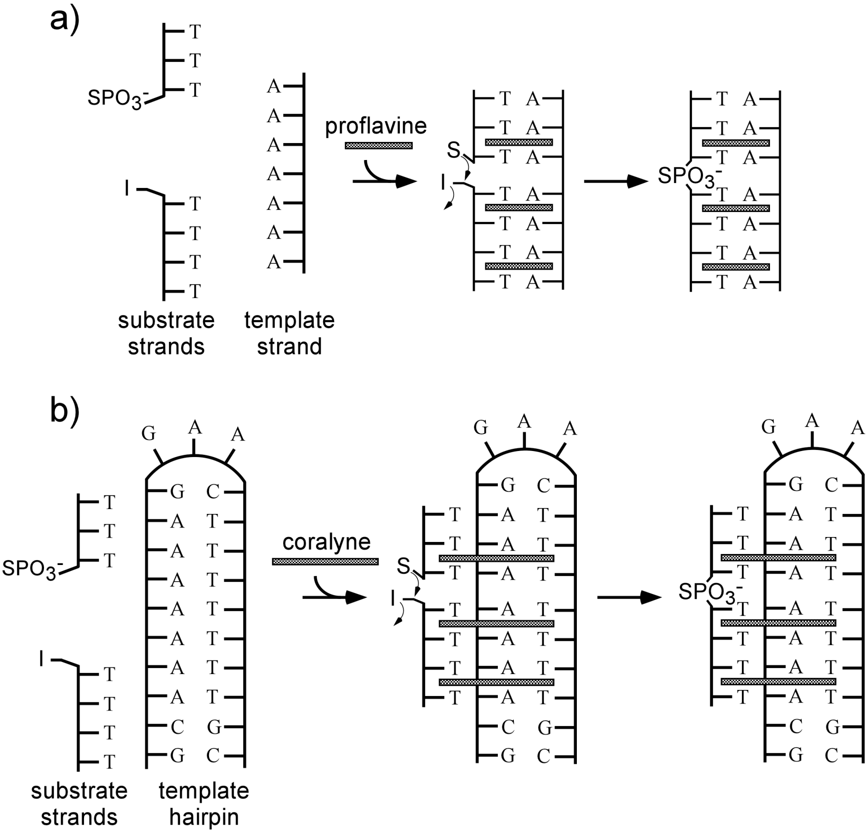


Figure 5. Experimental system used to explore the ability of intercalators to act as midwifes in a template directed ligation reaction including Watson–Crick base pairs or base triples. a) 3'-Phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$ substrate strands are too short to form a stable complex with the $(dA)_7$ template strand. The addition of proflavine promotes the assembly of a ligation-active complex, which leads to the formation of the $(dT)_7$ product within an internal phosphorothioate linkage. b) Coralyne facilitates the association of 3'-phosphorothioate-$(dT)_3$ and 5'-iodo-$(dT)_4$ substrate strands with a hairpin template to form a triplex ligation-active complex. Proflavine, which is specific for binding duplexes, is not active in promoting ligation in the presence of the hairpin template (41). (Adapted with permission from reference 41. Copyright 2007 Verlag Helvetica Chimica Acta.)

Intercalators could have also provided selective pressure toward the formation of a more uniform backbone. As discussed above, non-enzymatic template directed ligation reactions produce a mixture of 2′,5′ and 3′,5′ linkages, often with enrichment in 2′,5′ linkages (*65*). Recent work in our laboratory has demonstrated that proflavine has a 25-fold higher affinity for 2′,5′-linked versus 3′,5′-linked RNA (*66*). However, other intercalators, such as ethidium bromide, favor the 3′,5′-linked RNA duplexes (*66*). It can therefore be postulated that if a given midwife molecule has a preferential association with a 3′,5′-linked RNA backbone, then it might also promote the formation of this backbone linkage over the 2′,5′-linked alternative, particularly if reversible backbone linkage chemistries are employed (elaborated below).

Prebiotic molecular selection was certainly not restricted to multi-molecular complexes, such as the midwife stacks, but likely existed down to the molecular and metal ion levels. Sugars, with their tandem hydroxyl groups, are well suited for chelating cations. However, due to their variance in ionic radii and coordination geometries, different metal ions are best chelated by different sugars (*67*). Additionally, a given cation will not be chelated equally by all anomers of a sugar (e.g., pyranosyl versus furanosyl ribose) (*67*). Because the anomers of a sugar are in dynamic equilibrium, the relative population of anomers can be shifted in the presence of a cation, generating selection. For instance, the methylation of glucose results in 92% methyl pyranosides and 8% methyl furanosides under catalysis by HCl, which is close to the equilibrium distribution in water of 100% pyranosyl anomers (*28,68*). In contrast, methylation in the presence of $SrCl_2$ shifts the product distribution to 35% methyl furanosides as strontium is preferentially chelated by the furanosyl anomer of glucose (*68*). Given these experimental observations regarding the effects of metal ions on glycosylation reactions, it is conceivable that non-covalent interactions of salts, intercalators, or other small molecules played a significant role in the selective synthesis of nucleic acid monomers and polymers.

Thermodynamically Controlled Polymerization via Reversible Coupling Chemistries

Although molecular midwives could have facilitated the selection of the bases and the backbone length for the first proto-RNA polymers, the thermodynamic barriers to the formation of an RNA phosphodiester backbone still represent a formidable challenge to prebiotic synthesis. Artificial chemical activation to promote the formation of phosphodiester bonds, such as 1-methylimidazole or pyrophosphate activation of phosphates, results in low replication fidelity due to strand cyclization, base misincorporation, and premature product chain termination (*5,55,56*). These difficulties can be circumvented by using low energy, reversible initial bonding on the template, facilitating the selection of the most thermodynamically-favored base pairing and backbone conformation.

Reversible bond formation is utilized by living organisms during the first step of replication as a primary means of proofreading via the thermodynamic selection of the lowest energy base pair (i.e., correct Watson-Crick complement) (Scheme 1). The reversible condensation of the nucleoside triphosphate is followed by hydrolysis of the pyrophosphate product to kinetically trap the newly formed phosphodiester bond (Scheme 1). In emulating this two-step process, we maintain that both the initial step that exploits the stability of template association, and the subsequent reaction that traps the thermodynamically-favored product are necessary for high-fidelity replication.

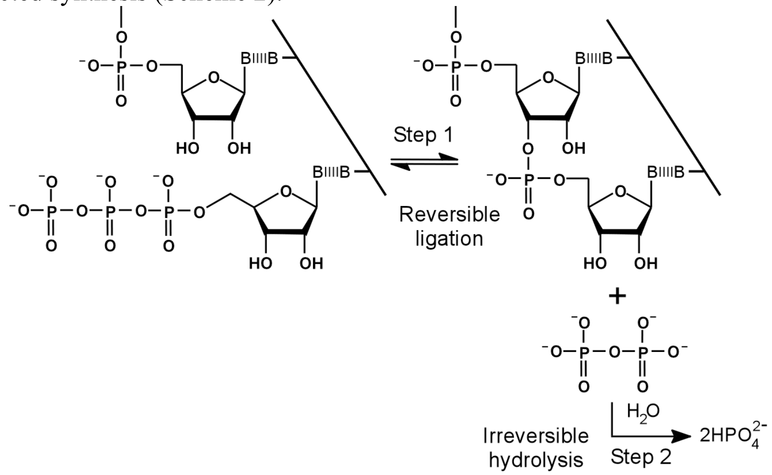
Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

Reductive Amination as a Test Case for the Utility of Reversible Backbone Linkages

To emulate the natural system, we have developed a nucleoside ligation chemistry employing reductive amination that indeed maintains the functional features of the two-step reversible coupling and irreversible trapping in a template-directed synthesis (Scheme 2). The first step of the ligation is the formation of an imine, a reversible reaction as indicated in the top panel of Scheme 2. Though reversible, the duplex formed by the association of the imine-linked polymer with the DNA template is highly favored (K_3, Scheme 2) (*69-71*). Surprisingly, only imine, and not the hydrated hemi-aminal intermediate, was observed in the associated complex contributing to the equilibrium. Therefore, even in water, where hydration is expected to contribute significantly to the equilibrium, the imine more significantly stabilizes the duplex over the hemi-aminal and functions as a good isosteric replacement of the phosphodiester bond.

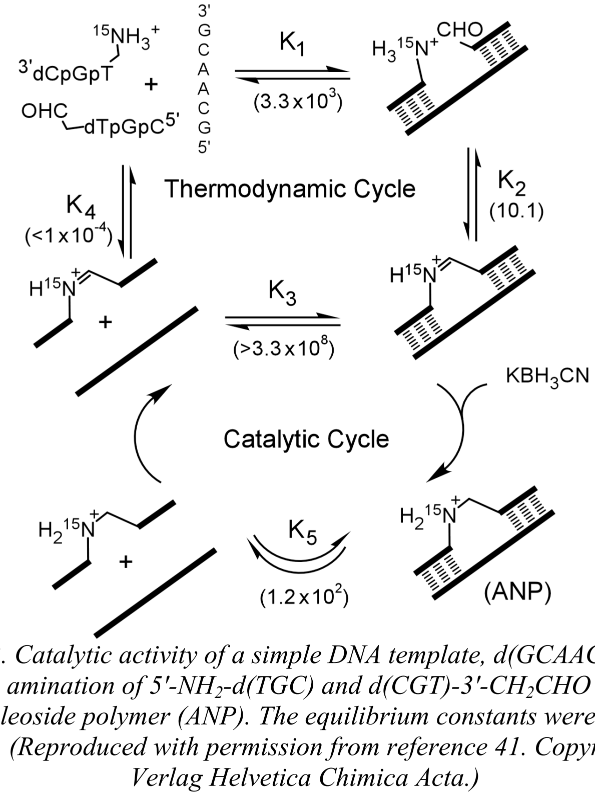
The irreversible second step is the chemical reduction of the intermediate imine to the amine to form amine-nucleoside polymers (ANPs) (lower panel in Scheme 2). This simple chemical change reduces duplex stability by a factor of nearly 10^6, which is attributed to the increased flexibility in the amine backbone arising from the removal of two conformationally restricted sp^2-hybridized centers (N and C) (*69-72*). This change in stability dramatically decreases product inhibition, which generally plagues template-directed oligonucleotide synthesis. Here, the product concentration must reach a 10^6-fold excess over the substrate concentration before it competes for catalytic sites along the template. More generally, these studies clearly establish that alternative backbone linkages can dramatically regulate the thermodynamics of template association, and this thermodynamic control can be exploited for catalytic cycles involving template-directed synthesis (Scheme 2).



Scheme 1. The biological synthesis of oligonucleotides proceeds via two steps, the first being a reversible ligation of the nucleotide triphosphate, and the second an irreversible hydrolysis of pyrophosphate. (Reproduced with permission from reference 41. Copyright 2007 Verlag Helvetica Chimica Acta.)

In order to demonstrate the utility of the imine/amine chemistry in a non-enzymatic template directed replication, mononucleotide and dinucleotide substrates were prepared with a 5′ amine and a 3′ aldehyde moiety (Figure 6). A general characteristic of bifunctional monomers is the intramolecular reaction to produce cyclic nucleotides and small cyclic oligonucleotides, severely reducing the active substrate concentration and diminishing product formation. While the T_1, T_NT and A_NT nucleotides can form intramolecular imines, no reaction was detected under mild reducing conditions in an aqueous environment, an observation attributed to the low imine concentration in the absence of a template (*73*). In the presence of a template and reducing conditions, these monomers condense to yield sequence and chain-length specific polymers.

As demonstrated by the product distribution of template-directed reductive polymerization of T_1 on a $(dAp)_8$ template in the presence of $NaBH_3CN$ (Figure 7), the polymerization proceeds through a step-growth mechanism. This polymer growth is a result of the reduced affinity of the amine products for the template strand versus the reactants or imine intermediates (Scheme 2). That is, as the association constant for T_1 imine dimer and the template is significantly greater than that of the product T_2, the rate of monomer reduction is greater than dimer, which again is greater than tetramer. Once the monomer supply is exhausted, T_2-T_2 condensation proceeds again at a higher rate than T_2-T_4 or T_4-T_4 couplings, and so on. The result is that no T_3, T_5, T_6, or T_7 intermediates are observed, and chain-length specific products are produced in high yield (Figure 7).



Scheme 2. Catalytic activity of a simple DNA template, d(GCAACG), on the reductive amination of 5′-NH_2-d(TGC) and d(CGT)-3′-CH_2CHO to form an amine-nucleoside polymer (ANP). The equilibrium constants were determined by NMR. (Reproduced with permission from reference 41. Copyright 2007 Verlag Helvetica Chimica Acta.)

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

The equilibrium control of replication inherent to the imine/amine system not only enables the synthesis of full-length products, but also favors high-fidelity sequence translation. To demonstrate this, a mixed sequence 32mer DNA template strand, 5′-dAAAAAT(AAAAAAAT)$_3$AA-3′, was replicated using T_1 and the amide-linked dimers T_NT and A_NT (Figure 6). The 32mer ANP complementary strand was synthesized as the sole product (>99%). Whether using a combination of T_1 and A_NT, or T_NT and A_NT as the substrates, no truncation products resulting from strand cyclization, nor misincorporated bases were detected by HPLC or mass spectrometry. There is now evidence that the transfer of sequence information can also be achieved in the reverse direction, from ANPs to DNA (*72*). These data clearly establish that the use of such two-stage coupling chemistries can achieve high-fidelity information transfer and template-defined chain length products, thereby eliminating cyclized side products. This example underscores the utility of equilibrium-controlled polymerization through reversible two-step non-enzymatic template-directed syntheses.

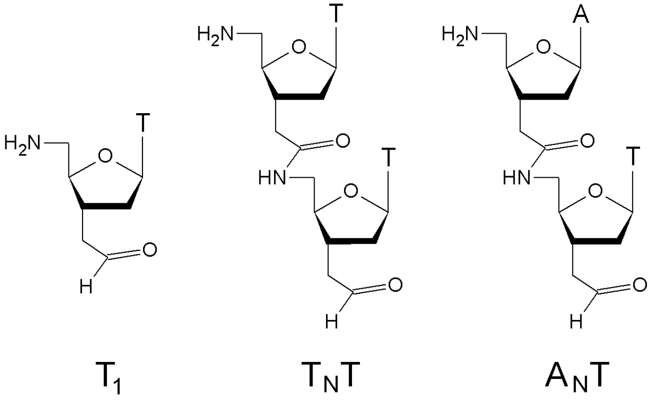


Figure 6. Bifunctional amine-nucleoside monomers used in the non-enzymatic template-directed replication of DNA templates.

Acetals as Low-Energy Predecessors to Phosphodiester Bond

The imine/amine backbone experiments established thermodynamically-controlled polymerization through initial reversible condensation on a template as a robust approach for achieving high fidelity transfer of sequence information without the aid of protein enzymes. Additionally, the ANP backbone is a good isosteric replacement of the phosphodiester backbone as it possesses six bonds from monomer to monomer and a tetrahedral geometry about the carbon of the amine linkage. As the ANP backbone is positively charged, it also satisfies the majority of the physical and chemical characteristics of phosphodiester-linked polymers, such as aqueous solubility, genotypic-phenotypic decoupling, and an extended backbone conformation. These physical and chemical characteristics would allow ANPs to fold into tertiary structures, replicate, and evolve.

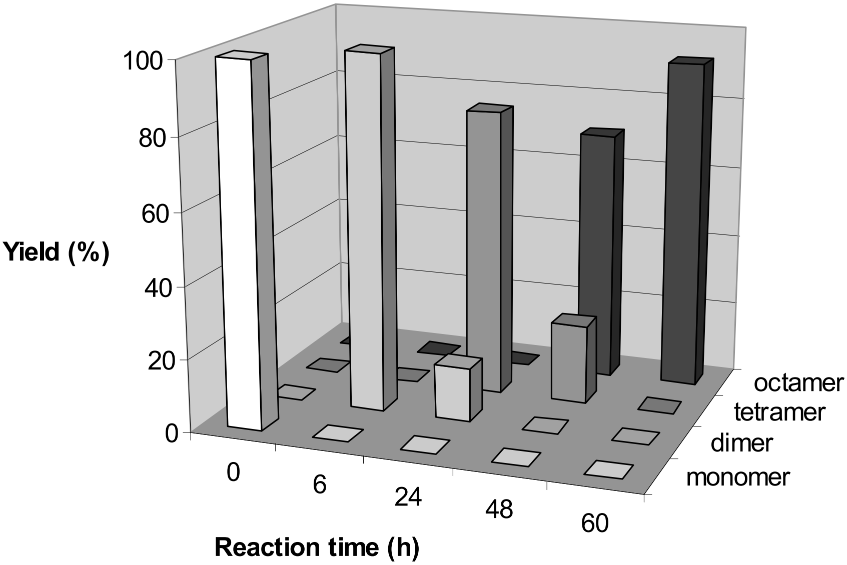


Figure 7. Product distribution upon reductive polymerization of T_1 on a $(dAp)_8$ template (41). The reactions contained 20 mol-equivalents of $NaBH_3CN$ in H_2O at 24°C. At 48h additional reducing agent was added to complete the reaction.

That said, metabolic interconversion of the amine backbone linkage of the ANPs would require oxidation, and thus would be a redox process rather than condensation/hydrolysis. Water remains a necessity for life and certainly life on Earth exploits condensation in the linkages of the nucleic acid monomers (phosphoester bonds), the amides of polypeptides, and the acetals of polysaccharides, and utilizes the reverse reaction of hydrolysis for metabolism. Therefore we have sought to extend the principles learned with the amine backbones to possible backbone chemistries that allow equilibrium control of template-directed polymerization through condensation and hydrolysis reactions.

Acetal-linked nucleic acids (*a*NAs) have several characteristics that make them interesting as possible progenitors of RNA via this approach (Figure 8) (*61*). First, acetals readily equilibrate in H_2O for two-step error correction during replication, and condensation is enthalpically favored over hydrolysis upon water removal (*74*). Second, the *a*NA backbone is structurally similar to the RNA backbone as both the acetal bond and phosphodiester bond have a tetrahedral geometry (Figure 8). Moreover, it is already known that as phosphodiester isosteres, *a*NAs can base pair with RNAs, (*75*). Third, acetal linkages can be functionalized to include charged groups that would maintain the chemical and physical properties of the nucleic acids as discussed above.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

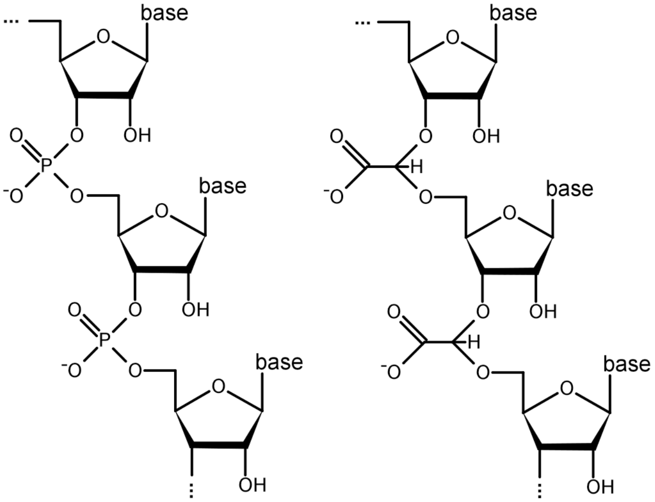


Figure 8. A comparison of the RNA (left) and the glyoxylate-acetal nucleic acid, or gaNA, (right) linkages. The phosphodiester linkage of RNA and the acetal linkage of gaNA possess a tetrahedral geometry and six bonds from monomer to monomer, resulting in similar helical properties of their duplexes (57). However, the acetal linkage is enthalpically favored over its hydrolysis product, and is more kinetically accessible than phosphodiester bonds.

We have proposed that glyoxylate (the ionized form of glyoxylic acid) could have preceded phosphate in proto-RNA backbones, forming glyoxylate-acetal nucleic acids (*ga*NAs) (Figure 8) (*50*). Glyoxylate is the smallest aldehyde molecule that is negatively charged under neutral conditions, and its prebiotic synthesis from glycolaldehyde has been demonstrated (*76*). The acetal bond formed by glyoxylate is a close structural and electrostatic analogue to the phosphodiester bond and a duplex consisting of glyoxylate-linked ribonucleosides is calculated to have helical properties very similar to RNA, suggesting the ability for *ga*NA and RNA strands to base pair (*50*). Additionally, we have shown that *ga*NA dinucleotides can be formed through simple heating-drying reactions starting from a neutral aqueous mixture of nucleosides, sodium glyoxylate, and magnesium chloride. These acetal bonds are quite stable, with no observable degradation after two weeks at room temperature in neutral aqueous solution (*50*). However, at elevated temperatures the glyoxylate linkages become more labile, providing a temperature regime under which the bonds are reversible and equilibrium controlled replication could take place. Therefore, the glyoxylate-linked backbone achieves a proto-RNA backbone that is structurally similar to RNA, but much more easily assembled and perhaps replicated.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

Evolution and the Emergence of the RNA World

Selective chemistries, such as supramolecular assemblies formed by bases and molecular midwives, and backbone formation through reversible coupling chemistries could have provided a route from the monomer world of the prebiotic chemical inventory to a proto-RNA world that was comprised of easily-assembled, thermodynamically favored polymers (Plate 2). Although molecular midwives would have promoted the synthesis of proto-RNA polymers with a particular backbone length and helix width, the chemical composition of these polymers may not have been homogeneous. We envision that at the dawn of the proto-RNA world, before the rise of catalytic RNAs, the oligonucleotide backbones would have been composed of a mixture of sugarsand low-energy linking molecules and perhaps bases that resembled those of current RNA (i.e., able to Watson-Crick base pair), but not of identical chemical structure. However, as catalytic proto-RNAs were selected through Darwinian evolution, proto-ribozymes that facilitated RNA synthesis (e.g., ribose synthesis, base synthesis, glycosylation, and phosphodiester bond formation, etc.) could have gradually promoted the emergence of RNA-like polymers that were more uniform, stable, functional, and more like RNA (Plate 2).

Acknowledgments

The research described was supported by the NSF (CHE-0404677, CHE-0739189), the NASA Exobiology Program (NNX08A014G), and DOE (ER15377).

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

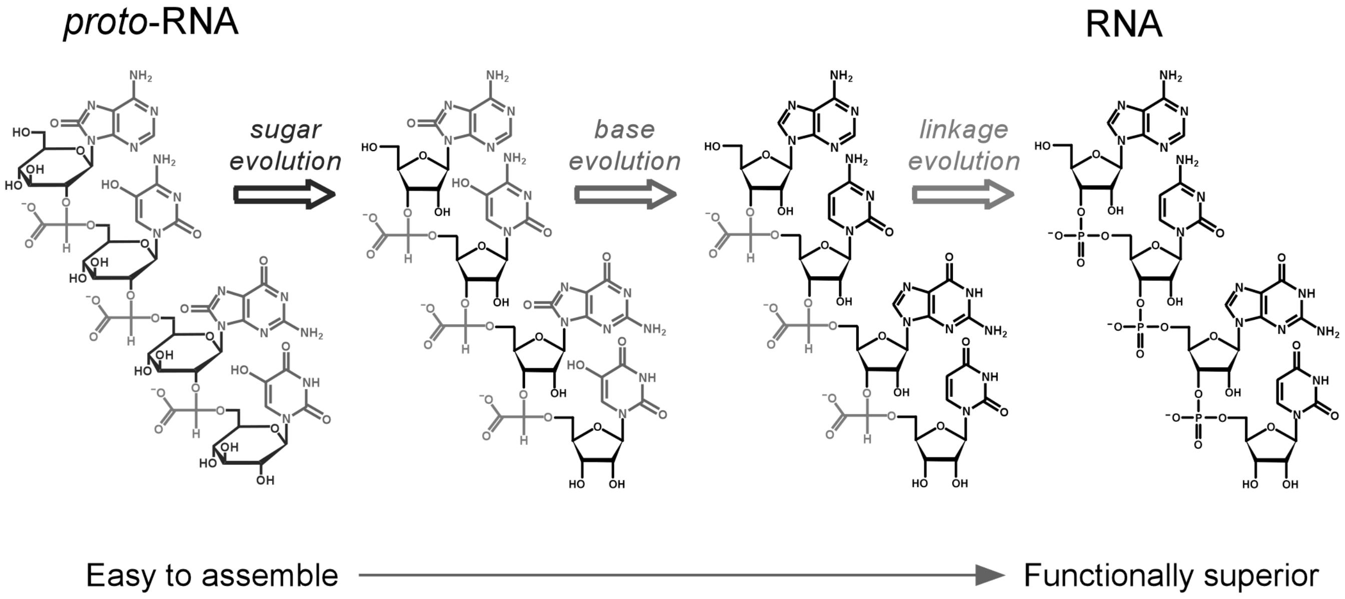

proto-RNA
sugar evolution
base evolution
linkage evolution
RNA
Easy to assemble
Functionally superior

Plate 2.(see color insert 2) The earliest proto-RNA polymers would have been composed of monomer units that were easy to assemble. Although the proto-RNA depicted here has a backbone comprised of a single type of sugar (blue) and a single linking molecule (green), in the absence of RNA or protein enzymes the earliest proto-RNAs were likely more heterogeneous. As proto-RNA sequences that promoted RNA synthesis were spontaneously generated and selected through Darwinian evolutionary processes, the populations of proto-RNAs would have become more homogeneous, stable, and functional, eventually leading to the catalyzed synthesis of RNA.

References

1. Woese, C. *The Genetic Code. The Molecular Basis for Genetic Expression;* Harper & Row: New York, 1967; pp 179–195.
2. Crick, F. H. C. *J. Mol. Biol.* **1968,** *38,* 367–379.
3. Orgel, L. E. *J. Mol. Biol.* **1968,** *38,* 381–393.
4. *The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World.;* Gesteland, R.; Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1999; p 709.
5. Joyce, G. F.; Orgel, L. E. In *The RNA World;* Gesteland, R. F.; Cech, T. R.; Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1999; pp 49–77.
6. Schwartz, A. W.; Chang, S. In *Life's Origin: The Beginnings of Biological Evolution;* Schopf, J. W., Ed.; University of California Press: Berkeley, CA, 2002; pp 46–77.
7. Hollis, J. M.; Jewell, P. R.; Lovas, F. J.; Remijan, A. *Astrophys. J.* **2004,** *613,* L45–L48.
8. Pizzarello, S. *Acc. Chem. Res.* **2006,** *39,* 231–237.
9. [cited 2008 24 February]; Available from: URL www.astrochemistry.net.
10. Kasting, J. F. *Science* **1993,** *259,* 920–926.
11. Rye, R.; Kuo, P. H.; Holland, H. D. *Nature* **1995,** *378,* 603–605.
12. Miller, S. L. *Science* **1953,** *117,* 528–529.
13. Oro, J. *Nature* **1961,** *191,* 1193–1194.
14. Reid, C.; Orgel, L. E. *Nature* **1967,** *216,* 455.
15. Reid, C.; Orgel, L. E. *Nature* **1967,** *216,* 936.
16. Miyakawa, S.; Yamanashi, H.; Kobayashi, K.; Cleaves, H. J.; Miller, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99,* 14628–14631.
17. Plankensteiner, K.; Reiner, H.; Schranz, B.; Rode, B. M. *Angew. Chem.* **2004,** *43,* 1886–1888.
18. Saladino, R.; Crestini, C.; Costanzo, G.; DiMauro, E. *Curr. Org. Chem.* **2004,** *8,* 1425–1443.
19. Cleaves, H. J.; Nelson, K. E.; Miller, S. L. *Naturwissenschaften* **2006,** *93,* 228–231.
20. Decker, P.; Schweer, H.; Pohlmann, R. *J. Chromatogr.* **1982,** *244,* 281–291.
21. Larralde, R.; Robertson, M. P.; Miller, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92,* 8158–8160.
22. Fuller, W. D.; Sanchez, R. A.; Orgel, L. E. *J. Mol. Evol.* **1972,** *1,* 249–257.
23. Fuller, W. D.; Sanchez, R. A.; Orgel, L. E. *J. Mol. Biol.* **1972,** *67,* 25–33.
24. Kolb, V. M.; Dworkin, J. P.; Miller, S. L. *J. Mol. Evol.* **1994,** *38,* 549–557.
25. Dworkin, J. P. Ph. D. thesis, University of California, San Diego, CA, 1997.
26. Bean, H. D.; Sheng, Y. H.; Collins, J. P.; Anet, F. A. L.; Leszczynski, J.; Hud, N. V. *J. Am. Chem. Soc.* **2007,** *129,* 9556–9557.
27. Angyal, S. J. *Aust. J. Chem.* **1968,** *21,* 2737–2746.
28. Angyal, S. J.; Pickles, V. A. *Aust. J. Chem.* **1972,** *25,* 1695–1710.
29. Angyal, S. J.; Pickles, V. A. *Aust. J. Chem.* **1972,** *25,* 1711–1718.
30. Bolli, M.; Micura, R.; Eschenmoser, A. *Chem. Biol.* **1997,** *4,* 309–320.
31. Eschenmoser, A. *Science* **1999,** *284,* 2118–2124.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

32. Schöning, K. U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000,** *290,* 1347–1351.
33. Eschenmoser, A. *Origins Life Evol. Biospheres* **2004,** *34,* 277–306.
34. Egli, M.; Pallan, P. S.; Pattanayek, R.; Wilds, C. J.; Lubini, P.; Minasov, G.; Dobler, M.; Leumann, C. J.; Eschenmoser, A. *J. Am. Chem. Soc.* **2006,** *128,* 10847–10856.
35. Sanchez, R. A.; Orgel, L. E. *J. Mol. Biol.* **1970,** *47,* 531–543.
36. Ingar, A.-A.; Luke, R. W. A.; Hayter, B. R.; Sutherland, J. D. *ChemBioChem* **2003,** *4,* 504–507.
37. Anastasi, C.; Crowe, M. A.; Sutherland, J. D. *J. Am. Chem. Soc.* **2007,** *129,* 24-25.
38. Crowe, M. A.; Sutherland, J. D. *ChemBioChem* **2006,** *7,* 951–956.
39. Wächtershäuser, G. *Proc. Natl. Acad. Sci. U. S. A.* **1988,** *85,* 1134–1135.
40. Zubay, G.; Mui, T. *Origins of life and evolution of the bioshpere* **2001,** *31,* 87–102.
41. Battersby, T. R.; Albalos, M.; Friesenhahn, M. J. *Chem. Biol.* **2007,** *14,* 525–531.
42. Ts'o, P. O. P.; Melvin, I. S.; Olson, A. C. *J. Am. Chem. Soc.* **1963,** *85,* 1289–1296.
43. Hud, N. V.; Jain, S. S.; Li, X. H.; Lynn, D. G. *Chem. Biodiversity* **2007,** *4,* 768–783.
44. Yang, Z. Y.; Hutter, D.; Sheng, P. P.; Sismour, A. M.; Benner, S. A. *Nucleic Acids Res.* **2006,** *34,* 6095–6101.
45. Krueger, A. T.; Lu, H. G.; Lee, A. H. F.; Kool, E. T. *Acc. Chem. Res.* **2007,** *40,* 141–150.
46. Mittapalli, G. K.; Osornio, Y. M.; Guerrero, M. A.; Reddy, K. R.; Krishnamurthy, R.; Eschenmoser, A. *Angew. Chem. Int. Ed.* **2007,** *46,* 2478–2484.
47. Mittapalli, G. K.; Reddy, K. R.; Xiong, H.; Munoz, O.; Han, B.; De Riccardis, F.; Krishnamurthy, R.; Eschenmoser, A. *Angew. Chem. Int. Ed.* **2007,** *46,* 2470–2477.
48. Westheimer, F. H. *Science* **1987,** *235,* 1173–1178.
49. Benner, S. A. *Acc. Chem. Res.* **2004,** *37,* 784–797.
50. Bean, H. D.; Anet, F. A. L.; Gould, I. R.; Hud, N. V. *Origins Life Evol. Biospheres* **2006,** *36,* 39–63.
51. Keefe, A. D.; Miller, S. L. *J. Mol. Evol.* **1995,** *41,* 693–702.
52. Costanzo, G.; Saladino, R.; Crestini, C.; Ciciriello, F.; Di Mauro, E. *J. Biol. Chem.* **2007,** *282,* 16729–16735.
53. Bryant, D. E.; Kee, T. P. *Chem. Commun.* **2006,** 2344–2346.
54. Pasek, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 853–858.
55. Kawamura, K.; Okamoto, F. *Viva Origino* **2001,** *29,* 162–167.
56. Miyakawa, S.; Joshi, P. C.; Gaffey, M. J.; Gonzalez-Toril, E.; Hyland, C.; Ross, T.; Rybij, K.; Ferris, J. P. *Origins Life Evol. Biospheres* **2006,** *36,* 343–361.
57. Orgel, L. E. In *Life's Origin: The Beginnings of Biological Evolution;* Schopf, J. W., Ed.; University of California Press: Berkeley, CA, 2002; pp 140–157.
58. Bernal, J. D. *P. Phys. Soc. Lond. A* **1949,** *62,* 537–558.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

59. Ferris, J. P.; Hill, A. R., Jr.; Liu, R.; Orgel, L. E. *Nature* **1996,** *381,* 59–61.
60. Orgel, L. *Origins Life Evol. Biospheres* **1998,** *28,* 227–234.
61. Hud, N. V.; Anet, F. A. L. *J. Theor. Biol.* **2000,** *205,* 543–562.
62. Ihmels, H.; Otto, D. In *Supermolecular Dye Chemistry;* Würthner, F., Ed.; Springer: Berlin, Germany, 2005; pp 161–204.
63. Chaires, J. B. *Biopolymers* **1997,** *44,* 201–215.
64. Jain, S. S.; Anet, F. A. L.; Stahle, C. J.; Hud, N. V. *Angew. Chem. Int. Ed.* **2004,** *43,* 2004–2008.
65. Ferris, J. P. In *Life's Origin: The Beginnings of Biological Evolution;* Schopf, J. W., Ed.; University of California Press: Berkeley, CA, 2002; pp 113–139.
66. Horowitz, E. D.; Hud, N. V. *J. Am. Chem. Soc.* **2006,** *128,* 15380–15381.
67. Angyal, S. J. *Chem. Soc. Rev.* **1980,** *9,* 415–428.
68. Parrish, F. W.; Angyal, S. J.; Evans, M. E.; Mills, J. A. *Carbohydr. Res.* **1975,** *45,* 73–83.
69. Zhan, Z.-Y. J.; Lynn, D. G. *J. Am. Chem. Soc.* **1997,** *119,* 12420–12421.
70. Gat, Y.; Lynn, D. G. *Biopolymers* **1998,** *48,* 19–28.
71. Luo, P. Z.; Leitzel, J. C.; Zhan, Z. Y. J.; Lynn, D. G. *J. Am. Chem. Soc.* **1998,** *120,* 3019–3031.
72. Ye, J.; Gat, Y.; Lynn, D. G. *Angew. Chem. Int. Ed.* **2000,** *39,* 3641–3643.
73. Li, X.; Zhan, Z.-Y. J.; Knipe, R.; Lynn, D. G. *J. Am. Chem. Soc.* **2002,** *124,* 746–747.
74. Wiberg, K. B.; Morgan, K. M.; Maltz, H. *J. Am. Chem. Soc.* **1994,** *116,* 11067–11077.
75. Matteucci, M. *Nucleosides Nucleotides* **1991,** *10,* 231–234.
76. Weber, A. L. *Origins Life Evol. Biospheres* **2001,** *31,* 71–86.

Downloaded by UNIV OF GUELPH LIBRARY on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch006

Chapter 7

Origin of Homochirality

Martin Klussmann,[1]* Donna G. Blackmond[2]

[1]Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany
[2]Department of Chemistry and Department of Chemical Engineering and Chemical Technology, Imperial College, London SW7 2AZ, UK

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

Chirality, the property of an object to be non-superimposable with its mirror image, is a common feature of biological molecules and has intrigued chemists for a long time. Of all the possible combinations, nature uses nearly exclusively L-amino acids and D-sugars and most other chiral molecules in enantiomerically pure form, too. This biological homochirality is regarded as a prerequisite of life as we know it, but how it evolved from a presumably racemic prebiotic Earth remains an open question. It is therefore linked to the general question of the origin of life on Earth itself. An overview is given on this topic focussing on recent experimental results, theoretical models and scenarios that might hold the clues to one of the biggest puzzles of science.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

The origin of life on Earth is one of the greatest and most fascinating puzzles of mankind. Scientists generally assume that it evolved in a stepwise process with increasing complexity, but the many questions of how exactly that happened remain essentially unanswered (*1*). One part of this big puzzle that is of particular interest for many chemists is the origin of biological homochirality.

The aim of this chapter is to give an overview on this topic by discussing a selection of observations, experimental results and scenarios, especially those related to amino acids. This emphasis allows the presentation of different models based on the same substance class in a closer context and also discussion of some more recent works. A more detailed survey of other models for the origin of biological homochirality can be found in review articles (*2-5*).

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

Homochirality – Definition and Importance

Chirality is a property of molecules or objects to be non-superimposable with their mirror-images. This is true for our right and left hands, from which the property derives its name. Many natural compounds are chiral, for example amino acids and sugars which are the building blocks of peptides, enzymes, RNA and DNA, the essential molecules of life. These key compounds occur in organisms almost exclusively in one enantiomeric form, amino acids in the L-, sugars in the D-form. This is referred to as *biological homochirality*. As a result, a vast number of other natural compounds, made with the help of chiral enzymes, also occurs enantiomerically pure in organisms.

This has important consequences for biochemical processes, for interactions of chiral compounds with organisms built from homochiral building blocks. The homochirality of receptors, enzymes and other key parts of an organism leads to a high degree of chiral discrimination, the ability to differentiate between enantiomers. Illustrative examples include smell and taste; for example, the terpenoid carvone in its (-)-form smells of mint whereas its enantiomer, the (+)-form, smells of caraway (*6*). The two enantiomers of a pharmaceutically active compound can display different effects as well, while one shows the desired effect, the other might be inactive or display different, possibly harmful activity (*7*, *8*).

More generally, the biochemical processes in any organism would fail to cooperate in the required way if the functional molecules (e.g. enzymes or RNA) were to consist of building blocks of random chirality. An enantiomer of an enzyme would still work in a mirror image manner; however, the random substitution of its amino acids by their enantiomers would eventually alter the enzyme's secondary or tertiary structure and thereby completely change or shut down its function. It is therefore generally accepted that homochirality is a basic requirement of life and its emergence is considered to be a fundamental step in prebiotic chemistry. The question remains if homochirality is a necessary precondition or an inevitable result of the emergence of life. Furthermore, is it the result of pure chance that all organisms today utilize L-amino acids and D-sugars; that is, could it as well have been the other way round? Or is there an intrinsic preference for this pair? Was it chance or determinism?

Despite these many questions being essentially unanswered, most of the discussions of the evolution of biological homochirality evoke a general stepwise model that can be experimentally supported. First, a process of *symmetry breaking* created a predominance of one enantiomeric form over the other. This imbalance was further enhanced by processes of *asymmetric amplification*, leading to high enantiomeric excesses in some compounds. These constituted either directly the building blocks for the evolution of life or indirectly by passing on their chirality through further chemical reactions.

Mirror Symmetry Breaking – is the Universe Really Racemic?

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

If synthesized in the laboratory from achiral or racemic starting materials, left and right-handed molecules of a compound are usually expected to form in equal amounts, a racemate. If enantiomerically pure or enriched substances are to be synthesized, chemists generally draw upon the natural pool of chiral molecules to provide enantiomerically pure building blocks or catalysts. Accordingly, one would expect the early universe to have been racemic along with all the compounds formed within it. Obviously this is not the case on Earth and in this section we will discuss some observations showing even the universe to be non-racemic in parts.

A possible explanation for mirror symmetry breaking in the universe comes from the observation of parity violation. This is a phenomenon that violates mirror symmetry in some nuclear processes involving weak interactions. It was first experimentally verified by showing that the radioactive β-decay of ^{60}Co predominantly produces particles with a left-handed spin (*9*). As a further consequence of parity violation, two enantiomers possess intrinsic differences in energy, thereby no longer being thermodynamically equal. This could explain why L-amino acids and D-sugars evolved to dominance over their enantiomers (*10*). But the calculated parity-violating energy differences (PVED) are of a very small magnitude, and while some calculations indeed suggest that the natural enantiomers are energetically favoured, others call this into question (*11*).

Because of these tiny and unclear PVED's, experimental consequences for chemistry are generally ruled out and have never been shown to exist beyond doubt (*12*). Parity violation is an accepted phenomenon in physics, but the question as to whether there is any connection with biomolecular homochirality on Earth remains open.

While the inanimate universe at a large might be racemic, there are observations of local imbalances. In some star-forming regions of the Orion Nebula, circular polarization (CP) has been found in the IR region; in some parts right-handed, in others left-handed CP dominates (*13*). Photochemical reactions with circularly polarized radiation have been shown to be able to produce small enantioselectivities, often by degrading one enantiomer faster than the other (*3*). Photolysis of racemic leucine with CP in the UV-region for example produced an ee of 2.5% after degradation of 75% of the material (*14*). It remains to be shown whether CP in the universe also exists in the UV region.

But there is more evidence that an enantiomerical imbalance exists in the universe. In two meteorites, the carbonaceous chondrites Murchison and

Murray, several organic compounds have been found including amino acids with significant enantiomeric excesses (ee) of up to 15% (*15-17*). Most of these do not occur naturally on Earth, eliminating the possibility of contamination with amino acids from Earth.

The percentage of low-molecular weight organic matter in meteorites like these is rather small, but accumulation of matter from space over billions of years could have resulted in significant amounts. It is assumed that extraterrestrial sources might have provided half or even more of all organic matter during the formation of the Earth (*18*, *19*).

Although the cause for these observations of imbalance in the universe are not fully understood, they might give us an echo from the time our solar system was formed. The universe as a whole might be racemic, but local imbalances have been found. Potentially, they provided the tipping of the balance in favor of L-amino acids and D-sugars when the Earth was formed.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

Asymmetric Amplification

Once a small imbalance between two enantiomers is established or given, this asymmetry can be further enhanced. Processes to turn low ee's into high ee's and possibly into the complete dominance of one enantiomer are called *Asymmetric Amplification*. Several examples are known, and many work fine under laboratory conditions but would fail under conditions presumed for the primordial Earth (*20*). Nevertheless, they might give an indication of the processes possible, indicating directions for further experiments on prebiotic chemistry scenarios.

Spontaneous Asymmetric Synthesis - Asymmetric Amplification in Autocatalysis

In 1953, Frank developed a mathematical model showing that spontaneous asymmetric synthesis is theoretically possible (*21*). If the chiral product of a catalytic reaction would act as a catalyst for its own formation and at the same time suppress the formation of its enantiomer, a basically enantiopure product could be formed from near-racemic starting materials. About forty years later, Soai and coworkers provided the first experimental proof for this concept of "asymmetric autocatalysis" with the alkylation of pyrimidyl aldehydes with dialkylzinc reagents (Figure 1) (*22*).

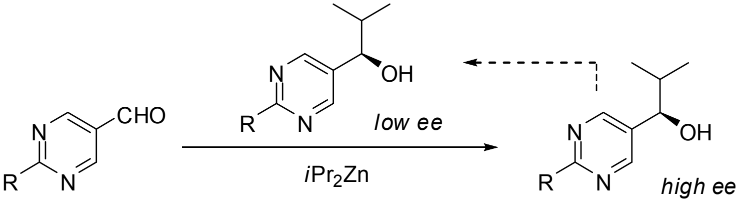


Figure 1. Asymmetric autocatalysis in the Soai reaction. R =H, Me, C≡CtBu.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

The product alcohol catalyses its own formation and the reaction shows spectacular asymmetric amplification. If small amounts of product with 5% ee are added at the beginning of the reaction, new product is formed with 55% ee. If this product is used as a catalyst in consecutive reactions, nearly enantiopure product is achieved after a few runs. Even starting with completely racemic material, the reaction product is generally produced non-racemic, with stochastically either one or the other enantiomer in excess (*23*). Other chiral compounds can direct the reaction towards the selective production of one particular enantiomer as well. About any form of chiral template has been shown to induce this effect, from helical hydrocarbons to chiral quartz crystals. The mechanism of this remarkable reaction was deduced with the help of kinetic studies and involves catalytically active homochiral dimers and inactive heterochiral ones (*24, 25*).

To date, this reaction remains the best investigated and most successful example of an asymmetric autocatalytic reaction in full accord with Frank's model. While its sensitivity to water and limitation to special substrates excludes any involvement in prebiotic chemistry, it clearly shows the practical possibility of spontaneous asymmetric synthesis, and ongoing research might provide us with new discoveries in the future.

Very recently, autocatalysis has been found in organocatalytic Mannich reactions with chiral amines that are themselves made by catalysis with the amino acid proline (*26*). So far, no asymmetric amplification was observed but the stereoselectivity of the reaction catalysed by the enantiopure product was still very high. Like other autocatalytic reactions known where the product formed is of lower ee than the initially present product-catalyst, consecutive cycles of autocatalysis will inevitably lead to an erosion of enantiomeric purity. Let us therefore look at other means of asymmetric amplification that can provide the reverse: from low to high enantiomeric purity.

Chiral Crystals and Phase Behaviour

The perfect shape of crystals reflects their inner regular construction from a highly ordered array of molecules. The structure of a crystal determines thermodynamic properties and can make the crystal as a whole chiral, even if it is built from achiral components. In this section, some special properties of crystals will be looked at which allow for mirror symmetry breaking or amplification, focussing on phase behaviour studies. Most models discussed here work by unequally distributing enantiomers into different phases. Thus, the net amount of both enantiomers stays constant but enrichment in one phase can provide a *local* environment of high asymmetry.

A simple scenario can be imagined for the case of conglomerates, substances that give a mixture of enantiopure crystals if their racemate is crystallized. Each single crystal contains only the enantiopure substance but in the sum of all crystals it will be racemic. If by wind or water a few of these crystals are physically separated from the others they could by chance contain more crystals of one particular enantiomer. Thus, mirror symmetry would have been broken locally if further transformations would only happen with this crop (*27*).

Even achiral substances like quartz and $NaClO_3$ can crystallize with a chiral crystal structure, but under normal circumstances equal amounts of mirror-image crystals will be obtained. But if crystallization is performed while stirring, continuous break-up and growth of the very first crystal will generate crystals of nearly uniform handedness (*28*). Even starting with a batch of racemic crystals, the solid phase can be driven towards chiral purity by utilising glass balls for an abrasion-grinding process that makes one enantiomeric form grow at the expense of the other (*29*, *30*). As sodium chlorate itself is achiral, the achieved homochirality would be lost if it is redissolved. But the same process could also be applied to a chiral amino acid derivative that crystallizes as a conglomerate; under racemising conditions in solution, the solid phase could be driven towards chiral purity (*31*). Additionally, processes of selective adsorption on chiral crystal surfaces are known that could have transferred the crystal-phase chirality to other compounds (*5*); for example, stereoselective adsorption of racemic amino acids can occur on chiral quartz or calcite faces, leading to small enantiomeric excesses in the adsorbed substance (*32*).

The majority of chiral substances do not crystallize as a conglomerate but belong to the class of so-called racemic compounds, which prefer to give racemic crystals with the two enantiomers arranged in alternating order (*33*). If they are crystallized with one enantiomer in excess, the racemic crystals will be mixed with a second type of crystal containing only the excess enantiomer. If to such a mixture of racemic and enantiopure crystals water or any solvent is added, so that saturated solution in equilibrium with excess of both crystal types is gained, the composition in the solution is fixed at a given temperature and pressure (Figure 2).

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

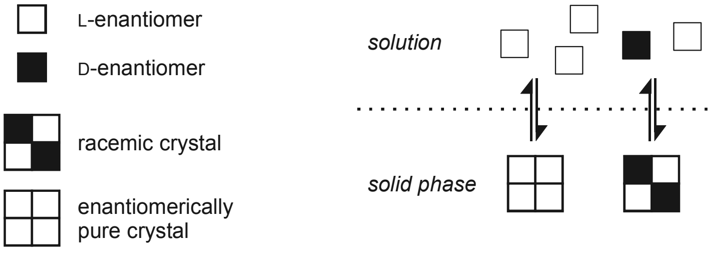


Figure 2. Common crystal forms of chiral substances and solid-solution equilibrium of an enantiomerically enriched substance crystallising as racemic compound.

This fixed composition results from the phase rule and is also called a eutectic point. The enantiomeric composition in solution can be higher or lower than initially employed, depending on the eutectic point. In the solid phase it will be the other way round, as no new enantiomers are made or consumed. For amino acids in water, the whole range from a low to high eutectic point is found: threonine is a conglomerate, which is always racemic at the eutectic point, whereas serine exhibits a eutectic point of >99% ee, being basically enantiomerically pure (*34*). Other amino acids exhibit high eutectic compositions, too, as shown in Figure 3.

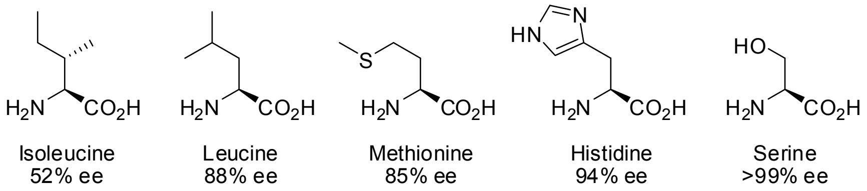


Figure 3. Enantiomeric composition of selected L-amino acids at the eutectic point in water at 25°C. Data taken from references 34, 35.

In contrast to the models discussed above for conglomerates, this time it is the solution phase not the solid that can become enantiomerically enriched, a possibility that had already been pointed out by Morowitz in 1969 (*36*). From the values shown in Figure 3, it seems at first that only a few selected amino acids allow for high ee's in solution. But further experiments with amino acids revealed ways to influence the eutectic composition. Valine for example has a medium eutectic of 47% ee in water at 25°C. Addition of fumaric acid, itself achiral, results in a rise of the eutectic to >99% ee (*37*). As it was shown, fumaric acid forms cocrystals with both racemic and enantiopure valine, strongly decreasing the solubility of racemic valine while not changing that of enantiopure valine much (Figure 4).

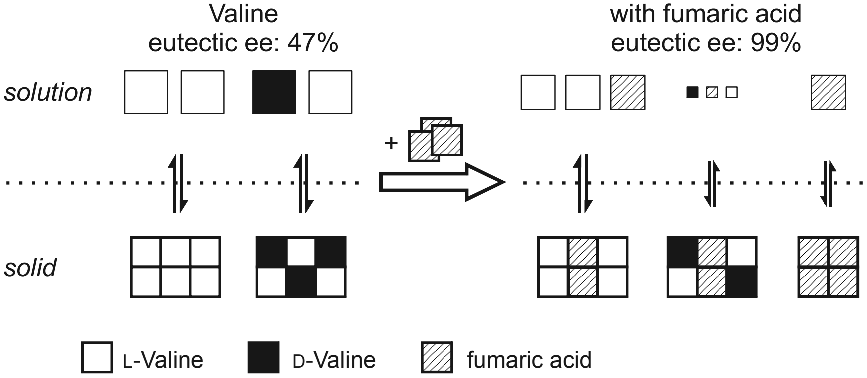


Figure 4: Enhancing the eutectic composition of valine by cocrystallization with an achiral additive. Adapted from reference 37.

Several cases of manipulating the eutectic ee by cocrystallization of amino acids with various achiral additives have been found, with ee enhancement as well as reduction. Such cocrystal formation appears to be very likely under prebiotic conditions that would have involved mixtures of different substances. Fumaric acid, for example, has also been found in the above mentioned meteorites containing amino acids.

In principle, the phase behaviour models discussed above apply for solid-gas phase systems, too (*33*, *38*). True equilibrium behaviour is less easily

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

investigated in this case, but the implications for sublimation processes, i.e. mass transfer, are clear and analogous to solid-solution systems. Accordingly, incomplete sublimation of amino acids in vacuum at elevated temperatures can lead to significant enantioenrichment in the sublimate. Sublimation of serine with an initial ee of 3% yielded sublimates of 68-92% ee (*39*), similar results have also been obtained with several other amino acids (*40*). Potentially, such processes could have taken place in the vacuum of space, leading to enantioenriched amino acids which reached the Earth *via* meteorites, providing a source of enantioenriched molecules. Serine is worthy of mention in this context as it also forms ionized clusters in the gas phase with a strong preference for homochirality and even the potential to selectively include other chiral compounds into these clusters (*41*). So far, these processes have not been shown to work under non-ionising conditions or in condensed phases, however (*42*).

On Earth itself, solid-solution phase systems of amino acids are not only much more likely than solid-gas phase systems, but have other interesting aspects as well. The fact that we are dealing with states of equilibrium means that asymmetric amplification could have been established over long periods of time, *via* cycles of rainfall and evaporation. Also, the equilibrium between solid and solution phase provides a “buffer” for slow racemization of the molecules; as long as the two different solid phases are maintained, regardless of their amount, the solution ee will remain at its eutectic value. The solution then sets the stage for another important phase in the evolution of biological homochirality, namely for chemical reactions. Thus, the amplified chirality could have been transferred to entirely new building blocks or improved asymmetric catalysts for such transformations could have been created.

Chirality Transfer and Preservation

Amino Acid Catalysis

Amino acids are not just building blocks of life but have been shown in recent times to be excellent asymmetric catalysts in synthetic chemistry. Foremost proline, but also other amino acids and derivatives catalyse many organic transformations with high enantiomeric excesses, amongst it the aldol reaction which provides access to another class of biological building blocks, carbohydrates (*43*). In general, these reactions are performed in organic solvents for synthetic purposes; the performance in water often suffers from reduced efficiency and stereoselectivity.

Isovaline, one of the amino acids present in meteorites in enantioenriched form, catalyses the synthesis of tetroses with ee’s up to 11% (*44*). This modest enantioselectivity is additionally hampered by the modest ee’s of isovaline present in the meteorite. But as discussed above, phase behaviour can provide for enantioenrichment in the solution phase and experiments showed that it can indeed be coupled to chirality transfer by catalysis in a single system. Enantiomerically pure L-serine for example catalyses an aldol reaction with 46% ee; if it is used with only 1% ee and the reaction performed in equilibrium with a solid phase of serine, the product is formed with identical stereoselectivity (Figure 5) (*34*).

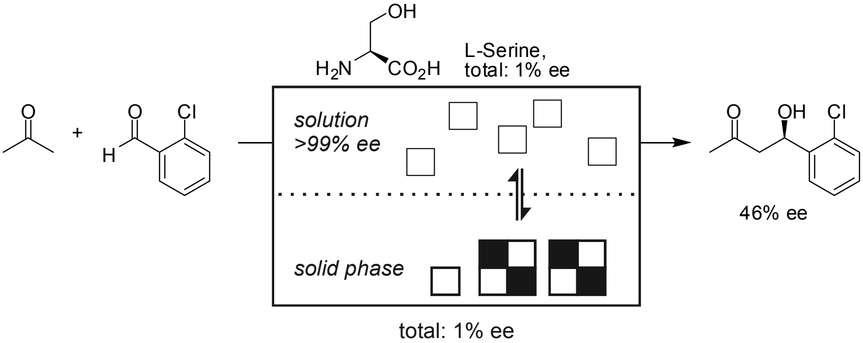


Figure 5: Asymmetric aldol reaction catalysed by L*-serine with a total ee of 1% in a solid-solution system giving rise to strong asymmetric amplification. Data from reference 34.*

The latter experiments were performed in DMF for reasons of efficiency but similar degrees of asymmetric amplification were achieved in water, too. The question that comes to mind is: is there a way to make improved catalysts?

Chirality in Macromolecules

Organisms do not use single amino acids as biocatalysts, but rather macromolecules comprised of them – enzymes – with nearly perfect selectivities in the functions they perform. But already the combination of just two amino acids to a dipeptide often creates a much better catalyst for the synthesis of carbohydrates under aqueous conditions (*45*, *46*). Alanine for example catalyses the formation of erythrose with 7% ee (*44*) and the dipeptide L-Ala-L-Ala with 33% ee (*46*). How exactly peptides evolved under prebiotic conditions is another topic of debate as peptide formation from amino acids in water is thermodynamically not favored, requiring special conditions and activations (*47*).

Yet apart from catalyst evolution, peptide formation might also hold some direct clues to the origin of homochirality. Macromolecules display some spectacular features resulting from their superstructure, such as cooperative effects of the monomers overruling a minority of “wrong” enantiomers in favor of one particular chiral backbone conformation (“majority rules”) (*48*). Peptides have been shown to be more stable towards hydrolysis when consisting of homochiral monomeric units and to be more reactive towards the uptake of a monomer of like chirality in polymerization reactions. Partial polymerization of leucine-*N*-carboxyanhydrides with an ee of 31% gave oligopeptides with 45% ee; partial hydrolysis resulted in the remaining polymer having an ee of 55% (*49*).

Synthetic macromolecules have also been shown to be capable of self-replication, an essential feature of nature’s macromolecules RNA and DNA and

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

of life itself. Many studies have been directed towards the creation of relatively simple self-replicating macromolecules that might serve as a model for the ancestors of modern RNA (*50*, *51*). The existence and evolution of such molecules is a central part of the “RNA-world”-hypothesis (*52*).

Self-replication has been shown to work with peptides as well. Ghadiri and coworkers created a homochiral 32-residue polypeptide that could stereoselectively replicate itself from a racemic pool of fragments (Figure 6) (*53*).

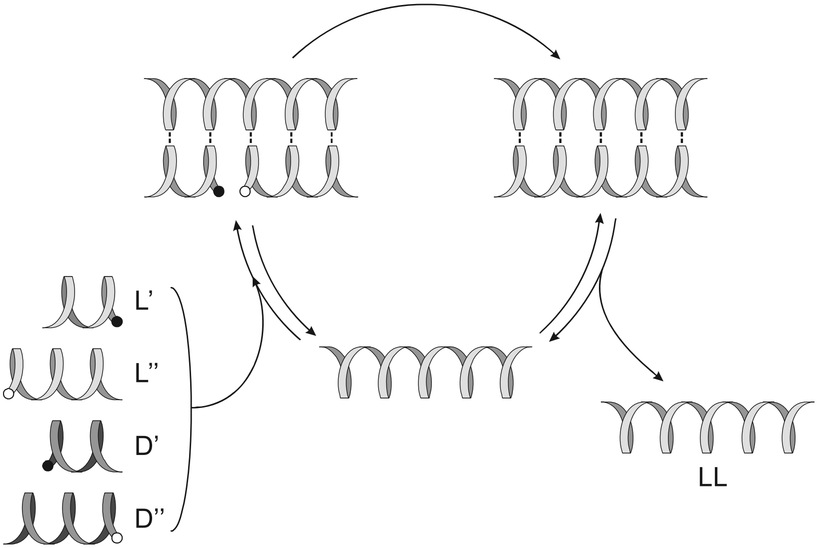


Figure 6. Schematic representation of chiroselective self-replication of the homochiral peptide LL. Adapted with permission from reference 53. Copyright 2001 Nature Publishing Group.

Moreover, it was also resistant towards small errors in the template chain: substitution of one monomeric unit by its enantiomer created a heterochiral template that nevertheless catalysed the formation of the homochiral peptide. The peptide’s homochirality would be preserved this way even if racemization of some of its building blocks occurred.

The remarkable features of macromolecules and the necessity to create functional macromolecules as an integral part of living organisms provides clues to how complete homochirality might have been achieved and stabilized. Homochirality is even suggested as an inevitable outcome during the evolution of such functional macromolecules (*54*). Still, processes of asymmetric amplification as discussed above could have helped to provide a pool of enantiomerically enriched building blocks that made the evolution of homochiral macromolecules much more feasible.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

Conclusions

Biological homochirality on Earth is a fact, but the puzzle of its evolution remains unsolved. Yet more and more separate discoveries add to the picture and some of the pieces seem to fit together to create possible scenarios. The universe as a whole might be racemic but some imbalances have been found like meteoritic amino acids with a slight enantiomeric excess. If these were collected by the prebiotic Earth in large amounts, a pool of building blocks would have existed for the first steps towards life and mirror symmetry would have been broken already. Through evaporation in tidal pools or lakes, solid amino acid crusts in equilibrium with saturated solutions could have enriched these pools in the excess enantiomers. Reactions in these solutions could have created peptides which, by their unique properties, could have further enhanced and preserved the achieved chirality, serving as templates for their own production or that of new products like carbohydrates.

Some pieces of the puzzle are clearly still missing and other scenarios exist that have not been explained here, like the "RNA-world" model suggesting that RNA was the first carrier of biological information and functional peptides evolved only later (*52*). But the more of such suggestions and experimental indications we have, the better we will be able to answer the question of how biological homochirality and finally life on Earth might have evolved.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

References

1. Luisi, P. L. *The Emergence of Life: From Chemical Orgins to Synthetic Biology;* Cambridge University Press: Cambridge, UK, 2006.
2. Bonner, W. A. *Orig. Life Evol. Biosph.* **1991,** *21,* 59–111.
3. Podlech, J. *Cell. Mol. Life Sci.* **2001,** *58,* 44–60.
4. Weissbuch, I.; Leiserowitz, L.; Lahav, M. *Topics in Current Chemistry* **2005,** *259,* 123–165.
5. Cintas, P. *Angew. Chem. Int. Ed. Engl.* **2002,** *41,* 1139–1145.
6. Leitereg, T. J.; Guadagni, D. G.; Harris, J.; Mon, T. R.; Teranishi, R. *Nature (London)* **1971,** *230,* 455–456.
7. Triggle, D. J. *Drug Discov. Today* **1997,** *2,* 138–147.
8. Crossley, R. *Tetrahedron* **1992,** *48,* 8155–8178.
9. Wu, C. S.; Ambler, E.; Hayward, R. W.; Hoppes, D. D.; Hudson, R. P. *Physical Review* **1957,** *105,* 1413–1415.
10. Yamagata, Y. *J. Theor. Biol.* **1966,** *11,* 495–498.
11. Wesendrup, R.; Laerdahl, J. K.; Compton, R. N.; Schwerdtfeger, P. *J. Phys. Chem. A* **2003,** *107,* 6668–6673.
12. Lahav, M.; Weissbuch, I.; Shavit, E.; Reiner, C.; Nicholson, G. J.; Schurig, V. *Orig. Life Evol. Biosph.* **2006,** *36,* 151–170.
13. Bailey, J.; Chrysostomou, A.; Hough, J. H.; Gledhill, T. M.; McCall, A.; Clark, S.; Ménard, F.; Tamura, M. *Science (Washington, D.C.)* **1998,** *281,* 672–674.
14. Flores, J. J.; Bonner, W. A.; Massey, G. A. *J. Am. Chem. Soc.* **1977,** *99,* 3622–3625.

15. Pizzarello, S.; Cronin, J. R. *Geochim. Cosmochim. Acta* **2000,** *64,* 329–338.
16. Pizzarello, S.; Zolensky, M.; Turk, K. A. *Geochim. Cosmochim. Acta* **2003,** *67,* 1589–1595.
17. Strasdeit, H. *ChemBioChem* **2005,** *6,* 801–803.
18. Ehrenfreund, P.; Irvine, W.; Becker, L.; Blank, J.; Brucato, J. R.; Colangeli, L.; Derenne, S.; Despois, D.; Dutrey, A.; Fraaije, H.; Lazcano, A.; Owen, T.; Robert, F. *Rep. Prog. Phys.* **2002,** *65,* 1427–1487.
19. Chyba, C.; Sagan, C. *Nature (London)* **1992,** *355,* 125–132.
20. Girard, C.; Kagan, H. B. *Angew. Chem. Int. Ed.* **1998,** *37,* 2922–2959.
21. Frank, F. C. *Biochim. Biophys. Acta* **1953,** *11,* 459–463.
22. Soai, K.; Shibata, T.; Morioka, H.; Choji, K. *Nature (London)* **1995,** *378,* 767–768.
23. Soai, K.; Shibata, T.; Sato, I. *Bull. Chem. Soc. Jpn.* **2004,** *77,* 1063–1073.
24. Blackmond, D. G.; McMillan, C. R.; Ramdeehul, S.; Schorm, A.; Brown, J. M. *J. Am. Chem. Soc.* **2001,** *123,* 10103–10104.
25. Blackmond, D. G. *Proc. Natl. Acad. Sci. USA* **2004,** *101,* 5732–5736.
26. Mauksch, M.; Tsogoeva, S. B.; Martynova, I. M.; Wei, S. *Angew. Chem. Int. Ed.* **2007,** *46,* 393–396.
27. Welch, C. J. *Chirality* **2001,** *13,* 425–427.
28. Kondepudi, D. K.; Kaufman, R. J.; Singh, N. *Science (Washington, D.C.)* **1990,** *250,* 975–976.
29. Viedma, C. *Phys. Rev. Lett.* **2005,** *94,* 065504.
30. Blackmond, D. G. *Chem. Eur. J.* **2007,** *13,* 3290–3295.
31. Noorduin, W. L.; Izumi, T.; Millemaggi, A.; Leeman, M.; Meekes, H.; Enckevort, W. J. P. V.; Kellogg, R. M.; Kaptein, B.; Vlieg, E.; Blackmond, D. G. *J. Am. Chem. Soc.* **2008,** *130,* 1158–1159.
32. Hazen, R. M.; Filley, T. R.; Goodfriend, G. A. *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 5487–5490.
33. Jacques, J.; Collet, A.; Wilen, S. H. *Enantiomers, Racemates and Resolution;* John Wiley: New York, 1981.
34. Klussmann, M.; Iwamura, H.; Mathew, S. P.; Wells jr., D. H.; Pandya, U.; Armstrong, A.; Blackmond, D. G. *Nature (London)* **2006,** *441,* 621–623.
35. Klussmann, M.; White, A. J. P.; Armstrong, A.; Blackmond, D. G. *Angew. Chem. Int. Ed.* **2006,** *45,* 7985–7989.
36. Morowitz, H. J. *J. Theor. Biol.* **1969,** *25,* 491–494.
37. Klussmann, M.; Izumi, T.; White, A. J. P.; Armstrong, A.; Blackmond, D. G. *J. Am. Chem. Soc.* **2007,** *129,* 7657–7660.
38. Blackmond, D. G.; Klussmann, M. *J. Chem. Soc. Chem. Commun.* **2007,** 3990–3996.
39. Perry, R. H.; Wu, C.; Nefliu, M.; Cooks, R. G. *J. Chem. Soc. Chem. Commun.* **2007,** 1071–1073.
40. Fletcher, S. P.; Jagt, R. B. C.; Feringa, B. L. *J. Chem. Soc. Chem. Commun.* **2007,** 2578–2580.
41. Takats, Z.; Nanita, S. C.; Cooks, R. G. *Angew. Chem. Int. Ed.* **2003,** *42,* 3521–3523.
42. Vandenbussche, S.; Vandenbussche, G.; Reisse, J.; Bartik, K. *Eur. J. Org. Chem.* **2006,** 3069–3073.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

43. Mukherjee, S.; Yang, J. W.; Hoffmann, S.; List, B. *Chem. Rev.* **2007,** *107,* 5471–5569.
44. Pizzarello, S.; Weber, A. L. *Science (Washington, D.C.)* **2004,** *303,* 1151.
45. Zou, W.; Ibrahem, I.; Dziedzic, P.; Sundén, H.; Córdova, A. *J. Chem. Soc. Chem. Commun.* **2005,** 4946–4948.
46. Weber, A. L.; Pizzarello, S. *Proc. Natl. Acad. Sci. USA* **2006,** *103,* 12713–12717.
47. Brack, A. *Chem. Biodiversity* **2007,** *4,* 665–679.
48. Green, M. M.; Park, J.-W.; Sato, T.; Teramoto, A.; Lifson, S.; Selinger, R. L. B.; Selinger, J. V. *Angew. Chem. Int. Ed.* **1999,** *38,* 3138–3154.
49. Blair, N. E.; Bonner, W. A. *Orig. Life Evol. Biosph.* **1981,** *11,* 331–335.
50. Robertson, A.; Sinclair, A. J.; Philp, D. *Chem. Soc. Rev.* **2000,** *29,* 141–152.
51. Patzke, V.; Kiedrowski, G. v. *ARKIVOC* **2007,** 293–310.
52. Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.* **2004,** *39,* 99–123.
53. Saghatelian, A.; Yokobayashi, Y.; Soltani, K.; Ghadiri, M. R. *Nature (London)* **2001,** *409,* 797–801.
54. Siegel, J. S. *Chirality* **1998,** *10,* 24–27.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch007

Downloaded by PENNSYLVANIA STATE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch008

Chapter 8

Surviving an Oxygen Atmosphere: DNA Damage and Repair

Cynthia J. Burrows

**Department of Chemistry, University of Utah, 315 S. 1400 East,
Salt Lake City, UT 84112-0850 USA**

As a consequence of life's coexistence with the reactive diradical O_2, cells have adapted biochemical defense mechanisms for protection from oxidative damage. Nevertheless, it is estimated that each cell's genomic DNA undergoes thousands of oxidative hits per day, and even more under conditions of stress. Unrepaired oxidative damage to DNA leads to mutations that underlie cancer, aging and neurological disease. Recent studies have helped unravel the oxidation chemistry of the DNA bases, and the myriad biochemical responses of DNA processing enzymes that battle against mutation. On the positive side, oxidative damage to nucleobases may accelerate the evolution of genomes and could have played a role in the ancestry of redox-active nucleoside cofactors as well as the adaptation of early life to changes in the environment.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

When cyanobacteria gained the function of converting H_2O to O_2 during the Great Oxygenation Event (~2.5 billion years ago), it was not clear how they survived the production of this toxic diradical. One theory proposes that during a "Snowball Earth" period, intense UV irradiation of glacial ice could have built up substantial levels of hydrogen peroxide. Subsequently, release of H_2O_2 during glacier melting, as well as its disproportionation to H_2O and O_2 might have helped drive evolution of oxygen-mediating enzymes (*1*). Eukaryotes emerged about 2 billion years ago in response to increasing levels of O_2 in the atmosphere. These organisms must have co-evolved numerous mechanisms to protect themselves from oxidative stress, that is, the over-production of reactive oxygen species (ROS) that would literally bleach critical biopolymers such as proteins and DNA.

In the present day, oxidative stress on the cell's genome stems from both endogenous and exogenous sources that include metabolic intermediates (superoxide and hydrogen peroxide) and products of inflammation (peroxynitrite and hypochlorous acid) (Figure 1). These ROS contribute to daily attack on DNA bases as well as on the sugar-phosphate backbone. Chemical reactions centered on the DNA bases do not generally break the DNA strand, as can happen when hydroxyl radical abstracts a hydrogen atom from the deoxyribose portion of the oligomer, but they are more likely to lead to mutations because base modification can disrupt the normal hydrogen bonding pattern of A:T and G:C base pairs.

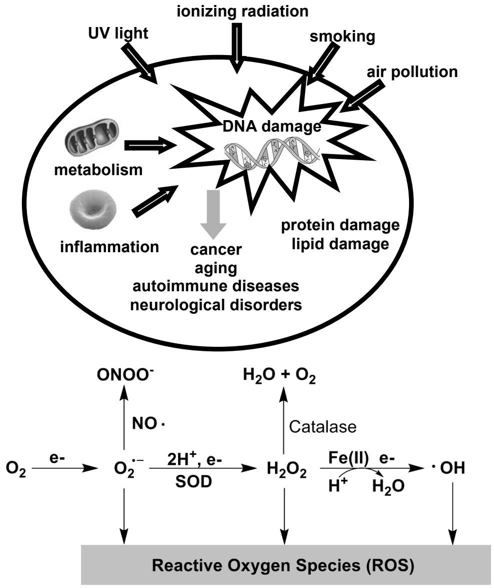

ionizing radiation
UV light
smoking
air pollution
DNA damage
metabolism
inflammation
cancer
aging
autoimmune diseases
neurological disorders
protein damage
lipid damage
ONOO⁻
H₂O + O₂
NO·
Catalase
O₂
e-
O₂·⁻
2H⁺, e-
SOD
H₂O₂
Fe(II) e-
H⁺ H₂O
·OH
Reactive Oxygen Species (ROS)

Figure 1.(see color insert 3) Reactive oxygen species, endogenously or exogenously produced, are major contributors to DNA damage leading to disease.

Downloaded by PENNSYLVANIA STATE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch008

Several mechanisms have evolved to seek out and repair changes to the DNA bases, hopefully before replication has sealed in a mutation (Figure 2). In a few cases, the integrity of the base can be restored by direct reversal of the damage. More commonly, minor changes are repaired by the base excision repair (BER) pathway that cleaves the glycosidic bond, releasing the damaged base (*2*). Further processing is required to remove the deoxyribose so that the appropriate nucleotide can be reinserted with a polymerase, and the ends are then sealed with a ligase. The BER pathway is particularly relevant to oxidative damage, because most reactions involve the addition of only one or two new oxygen atoms to the purine or pyrimidine base. For successively bulkier and more complex damages such as cross-links to other DNA strands or to proteins, more complicated pathways are involved. For example, in nucleotide excision repair (NER), a whole segment of DNA surrounding the lesion is clipped out and then resynthesized using the undamaged strand as a template. For more extensive damage, recombinational repair (not shown) may recruit the other copy of the chromosome. So dependent are organisms on DNA repair that several disease states have been associated with faulty or inadequate repair due to mutant enzymes (*3*).

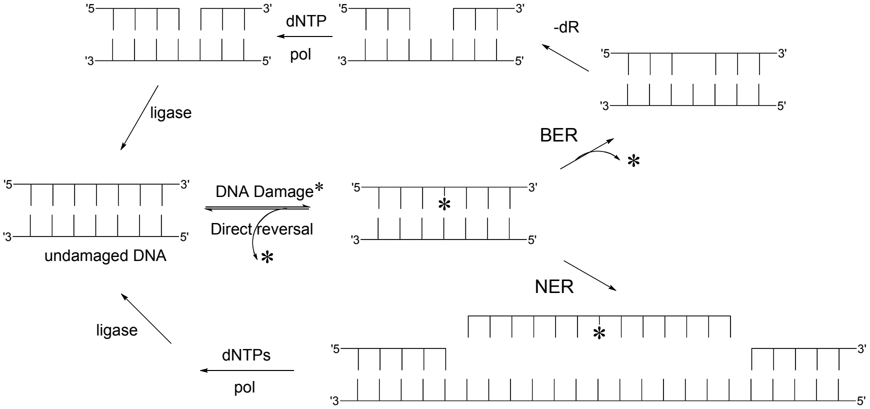


Figure 2. Three of the DNA repair pathways. Direct reversal of damage is rare. Base excision repair (top loop) in which the damaged base is cleaved from the ribose, then the ribose is removed leaving a gap filled in by a polymerase (pol) and religated with a ligase. Nucleotide excision repair (bottom loop) is similar to BER except a longer fragment surrounding the damage is removed.

Daily damage to the human genome is summarized in Figure 3 (*4*); depurination resulting from hydrolytic cleavage of the glycosidic bond is the most common outcome, and the non-instructive abasic site must be repaired. Fortunately, most DNA polymerases pause at abasic sites, permitting repair enzymes to be recruited before a faulty copy of the DNA template is synthesized. Another frequent damage type is deamination of cytosine yielding uracil, which would result in a C-to-T mutation if left unrepaired. Other modifications include formation of pyrimidine-pyrimidine dimers such as the T<>T cyclobutane dimer whose occurrence is highly dependent on exposure to UV light; indeed, it can easily climb to thousands of events per cell per day.

Downloaded by PENNSYLVANIA STATE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch008

Alkylation damage has a number of sources, many of them due to exogenous exposure to inhaled or dietary toxins, but S-adenosylmethionine-mediated (SAM) reactions that normally methylate C, can also erroneously methylate other bases, accounting for additional damage. Alkylation of N7 or N3 of purines enhances the lability of the glycosidic bond leading to abasic sites, therefore requiring repair; alkylation at other sites such as O6 of G can also be mutagenic.

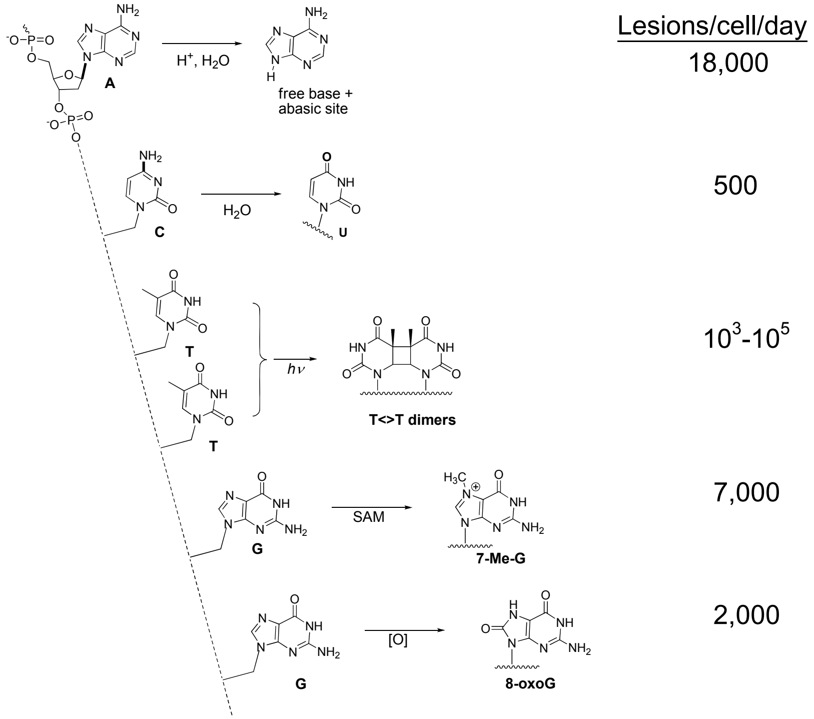


Figure 3. Structures of common DNA lesions and their frequency of occurrence.

Key among the detrimental damages to DNA bases are the oxidative transformations that occur principally on guanine, the most readily oxidized of the four bases. Measurement of 7,8-dihydro-8-oxo-2'-deoxyguanosine (8-oxoG), a biomarker of oxidative stress, has been intensively investigated, and the current estimate of this lesion under normal cellular conditions is approximately 1 in 10^6-10^7 guanosines (*5*). This means that each cell sees on average more than a thousand oxidized guanines per day. 8-OxoG is not the only product of oxidative damage to the bases; all four bases form adducts with hydroxyl radical, and G and A in particular form cyclic adducts with lipid oxidation products such as malondialdehyde.

Multiple pathways lead to the guanine oxidation product 8-oxoG (Figure 4) including one-electron oxidation of G, addition of hydroxyl radical (from Fenton reactions or ionizing radiation) and reduction of 8-hydroperoxyG, a proposed product of singlet oxygen addition to G, by cellular thiols. Other products of guanine oxidation include imidazolone (dIz) and formamidopyrimidine (Fapy-

Downloaded by PENNSYLVANIA STATE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch008

dG) derivatives. Yet more G oxidation products are produced as secondary oxidation products because of the much lower oxidation potential of 8-oxoG compared to the four natural DNA bases. Our laboratory has investigated the formation of hydantoin products from 8-oxoG oxidation; these comprise two diastereomers of spiroiminodihydantoin (Sp) as well as guanidinohydantoin (Gh) (Figure 5) (*6,7*). The latter's two diastereomers rapidly interconvert and cannot be studied separately. These products are formed in varying amounts depending on the medium and context; formation of Sp is disfavored in duplex DNA or at pH<6, but it is otherwise an abundant product of G and 8-oxoG oxidation.

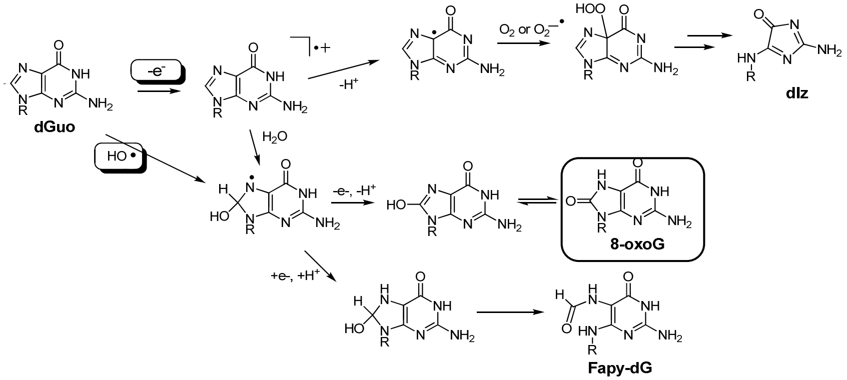


Figure 4. Pathways leading to 8-oxoG include ionizing radiation (HO˙) and one-electron oxidation.

Investigation of the mechanism of oxidation of G to form the hydantoin lesions was conducted with isotopically labeled G (*8*). Using ^{13}C, ^{15}N and ^{18}O heavy isotopes, we were able to determine that the oxygen atom incorporated at C8 was derived from O_2, while the oxygen atom introduced at C5 originated with H_2O. In this mechanism, water serves as a weak but abundant nucleophile adding to an oxidized quinone form of 8-oxoG. The significance of this study is that it provided us with the broader concept that oxidation of G leads to the formation of an electrophilic species (OG^{ox} in Figure 5), that could be trapped by any number of cellular nucleophiles. Examples of less abundant but better nucleophiles in the cell include the tetraamine spermine and the plethora of DNA-binding proteins which tend to be rich in lysine. Thus, oxidation of G can also trigger DNA-protein covalent cross-linking (*9*).

Downloaded by PENNSYLVANIA STATE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch008

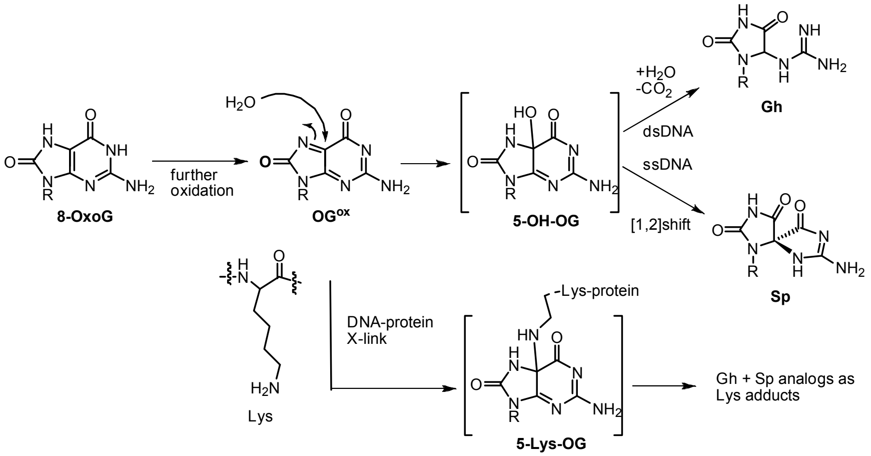


Figure 5. Formation of hydantoin products from further oxidation of 8-oxoG.

8-OxoG is weakly mutagenic in an *E. coli* assay leading to G-to-T transversion mutations. The molecular basis for this mutation is outlined in Figure 6 (*10*). Although the oxidative damage process that inserts an oxygen atom at C8 of G does not appear to alter the base-pairing properties of G, it does alter the preferred orientation of the base with respect to the glycosidic bond. Thus, 8-oxoG has a higher propensity to rotate into a *syn* conformation rather than the normal *anti* conformation. If it does so during replication, the 8-oxoG base presents a different hydrogen-bonding pattern compared to G, and its complement is A rather than C. If A is misincorporated into the strand opposite 8-oxoG, then there is a chance it will persist until the A-containing strand is replicated, in which case, a T would be placed opposite, in the original position of the G that underwent oxidation.

In contrast, the hydantoins are 99% mutagenic, mispairing with A or G and leading to G-to-T and G-to-C mutations, both of which are common outcomes of oxidative stress (*11*). The molecular basis of mutation via Gh and Sp is less well established. Presumably, the hydantoin ring resembles a T, albeit a distorted and ring-contracted one, so it may base pair with A, though not as well as the six-membered ring pyrimidine T. Pairing with G might be accomplished via shifting of the hydantoin ring toward the major groove to form a wobble pair; this would better accommodate the sp^3 carbon of the ring and the appended guanidinium (in Gh) or spirocycle (in Sp). (See Figure 6B.) When the hydantoin lesion mispairs with G, another round of replication would replace it with C; overall this would then be a G-to-C transversion mutation.

Downloaded by PENNSYLVANIA STATE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch008

Downloaded by PENNSYLVANIA STATE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch008

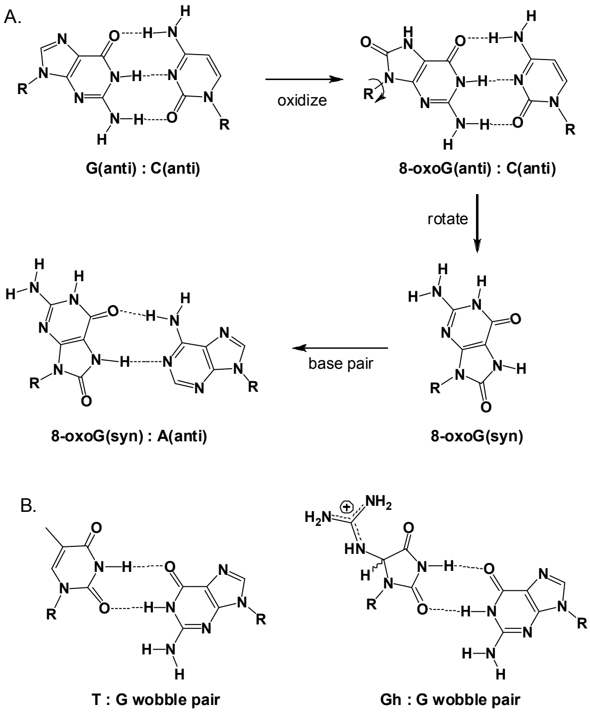


Figure 6. A. Structural basis of G-to-T transversion mutations via the 8-oxoG:A mispair. B. G-to-C mutations might occur via Gh formation that pairs with G, by analogy to T:G wobble pairs.

The high rate of mutagenicity observed for the hydantoin lesions Sp and Gh is due to their severely altered structures, compared to G, and their consequent inability to base pair correctly with C. The good news is that Sp and Gh can be recognized and excised by the base excision repair mechanism using a glycosylase called hNEIL1, the human homolog to bacterial Endonuclease VIII (Nei). In fact, we have found that the hydantoins are more than 100-fold better substrates than any other DNA lesion studied so far (*12*). Thus, the action of hNEIL1 on a Gh:C or a Sp:C base pair can correct the oxidative damage before mutation. Unfortunately, if replication has already occurred leading to misinsertion opposite the hydantoin (Gh:G, for example), the action of hNEIL1 will serve only to seal in the mutation by removing Gh and allowing its replacement with a new base in that position.

In the timeframe of a human lifespan, we think of mutation as a bad thing—a genetic component of aging and disease. However, without mutation, there would be no adaptation of species to changing environments, *i.e.* no evolution. Mutations can occur randomly due to the constant but low level of errors made by polymerases during replication. Alternatively, mutations can be accelerated due to DNA damage such as oxidation. Presumably, the Great Oxygenation Event of 2.5 billion years ago was responsible for some of the evolutionary pressure to generate proteins that could handle oxidative stress: superoxide dismutase and catalase to reduce the levels of ROS, and DNA repair enzymes to handle the oxidative damage to the genome.

One puzzle that has not been solved about the early evolution of life in an oxygen atmosphere is the question of redox cofactors for metabolic enzymes. The "RNA World" hypothesis suggests that ancient life evolved from the catalytic chemistry of RNA oligomers capable of basic functions like cleavage, ligation, and replication (*13*). A primitive form of metabolism was also needed, but the modern redox cofactors NADH and $FADH_2$ were not available. The persistence of the dinucleotide redox cofactors in evolutionary history has strengthened the argument that life evolved from RNA-based chemistry (*14*), but where did these cofactors come from? How did early RNA catalyze redox reactions without them?

We hypothesize that oxidized purines may have fulfilled this role in an earlier time. The 2-electron oxidation of 8-oxoG is remarkably similar to that of dihydroflavin, and an even closer parallel exists with the redox chemistry of pterins (Figure 7). Biosynthetically, pterins and flavins are derived from guanosine through a long series of transformations, but the current biosynthesis is presumably a highly evolved pathway that originated from a rudimentary derivative of G that was capable of redox chemistry. Perhaps that was 8-oxoG? In a single chemical step from G to 8-oxoG, the one-electron redox potential of the purine is lowered by 600 mV ($E_{1/2}$ of G is 1.3 V vs. NHE; see Figure 7 for others). Although this is still higher than pterins and flavins, it is nevertheless a huge step in the direction of forming a redox-active ribonucleoside, potentially capable of assisting primitive proteins to develop an oxygen-based metabolism for evolving organisms. We cannot travel back in time to observe early evolution, but current experiments may lend credence to this hypothesis.

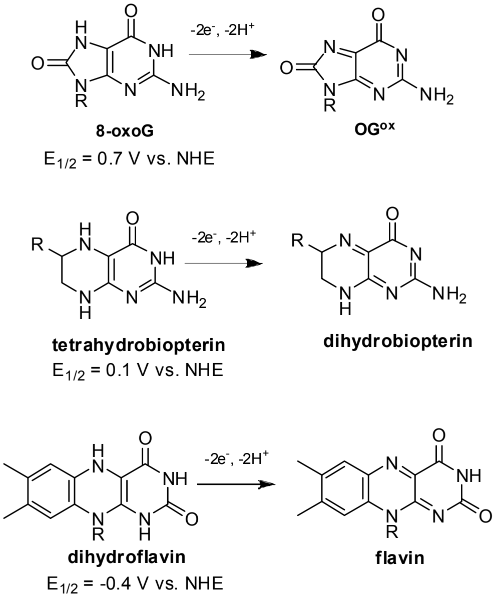


Figure 7. Analogy between redox chemistry of 8-oxoG and pterins and flavins.

Downloaded by PENNSYLVANIA STATE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch008

Oxidative DNA damage and repair processes are like two ends of a seesaw—when damage increases, repair must be upregulated to keep the genomic mutation rate in balance. Otherwise, oxidative stress could become catastrophic for the cell, or at a minimum, may cause permanent changes in the genome. On the positive side, the evolution of the species depends on mutation, and indeed the evolution of life as we know it may have depended on oxidative purine chemistry to utilize the oxygen atmosphere to its advantage.

Acknowledgements

Various aspects of this work from the author's laboratory were funded by grants from NIH (CA090689) or NSF (CHE0809483); their support is gratefully acknowledged. The contributions of numerous coworkers, whose names appear in the references, have been monumental.

Downloaded by PENNSYLVANIA STATE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch008

References

1. Liang, M.-C.; Hartman, H.; Kopp, R. E.; Kirschvink, J. L.; Yung, Y. L. Production of hydrogen peroxide in the atmosphere of a Snowball Earth and the origin of oxygenic photosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103,* 18896–18899.
2. David, S. S.; O'Shea, V. L.; Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **2007,** *447,* 941.
3. Hoeijmakers, J. H. J. Genome maintenance mechanisms for preventing cancer. *Nature* **2001,** *411,* 366–374.
4. Friedberg, E.; Walker, G. C.; Seide, W.; Wood, R. D.; Schultz, R. A.; Ellenberger, T. A. *DNA Repair and Mutagenesis;* ASM Press: Washington, DC, 2006.
5. Cadet, J.; Douki, T.; Ravanat, J.-L. Oxidatively generated damage to the guanine moiety of DNA: Mechanistic aspects and formation in cells. *Acc. Chem. Res.* **2008,** *41,* 1075–1083.
6. Luo, W.; Muller, J. G.; Rachlin, E. M.; Burrows, C. J. Characterization of spiroiminodihydantoin as a product of one-electron oxidation of 8-oxo-7,8-dihydroguanosine. *Org. Lett.* **2000,** *2,* 613–617.
7. Luo, W.; Muller, J. G.; Rachlin, E. M.; Burrows, C. J. Characterization of hydantoin products from one-electron oxidation of 8-oxo-7,8-dihydroguanosine in a nucleoside model. *Chem. Res. Toxicol.* **2001,** *14,* 927–938.
8. Ye, Y.; Muller, J. G.; Luo, W.; Mayne, C. L.; Shallop, A. J.; Jones, R. A.; Burrows, C. J. Formation of ^{13}C, ^{15}N, and ^{18}O-labeled guanidinohydantoin from guanosine oxidation with singlet oxygen. Implications for structure and mechanism. *J. Am. Chem. Soc.* **2003,** *125,* 13926–13927.
9. Xu, X.; Muller, J. G.; Ye, Y.; Burrows, C. J. DNA-Protein cross-links between guanine and lysine depend on the mechanism of oxidation for formation of C5 Vs C8 guanosine adducts. *J. Am. Chem. Soc.* **2008,** *130,* 703–709.

10. Neeley, W. L.; Essigmann, J. M. Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chem. Res. Toxicol.* **2006,** *19,* 491–505.
11. Henderson, P. T.; Delaney, J. C.; Muller, J. G.; Neeley, W. L.; Tannenbaum, S. R.; Burrows, C. J.; Essigmann, J. M. The hydantoin lesions formed from oxidation of 7,8-dihydro-8-oxoguanine are potent sources of replication errors *in vivo*. *Biochemistry* **2003,** *42,* 9257–9262.
12. Krishnamurthy, N.; Zhao, X.; Burrows, C. J.; David, S. S. Superior removal of hydantoin lesions relative to other oxidized bases by the human DNA glycosylase hNEIL1. *Biochemistry* **2008,** *47,* 7137–7146.
13. *The RNA World;* Gesteland, R. F.; Atkins, J. F.; Eds.; Cold Spring Harbor Press: Plainview, NY, 1993.
14. White, H. B. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **1976,** *7,* 101–104.

Downloaded by PENNSYLVANIA STATE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch008

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

Chapter 9

Isotopes Illuminate Chemical Change: Boron Isotope *p*H Proxy

N. Gary Hemming[1,2]

[1]School of Earth and Environmental Sciences, Queens College, Flushing, NY 11367
[2]Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964

A driving force for the evolution of life on Earth is adaptation to environmental change. Basic chemical principles are used by geochemists to study the evolution of the Earth system, including ancient climate. Because there is no way to obtain climate information directly, geochemists rely on proxies, and isotopes have been particularly useful tools. Boron isotopes are promising as a proxy for ancient ocean *p*H and for understanding natural variations in atmospheric CO_2 concentrations, due to the tight coupling between the atmosphere and surface ocean. Proxies for CO_2, combined with temperature proxies, are important for understanding the natural relationships between global warming and atmospheric CO_2, and thus will aid models that seek to predict future warming. The development of any proxy requires an understanding of fundamental chemical principles including atomic structure, vibrational energies, aqueous speciation, and isotopic fractionation, all within a dynamic system that is perturbed by biological influences.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Introduction

Presented here are the basic chemical principles behind the development of a new tool for understanding changes in ocean chemistry in the past. The boron isotope paleo-*p*H proxy allows a measure of the *p*H of ancient oceans, and because there is a coupling between the surface ocean and the atmosphere, it also allows an estimation of atmospheric CO_2 concentrations in the past. This is essential information because the better we understand the natural fluctuations and controls on Earth's carbon budget, the better we can understand, predict, and ameliorate the consequences of the post-industrial increase in atmospheric CO_2. Humans are having a profound impact on Earth's chemistry, and the recent increase in atmospheric CO_2 concentration is alarming. David Keeling began monitoring atmospheric CO_2 concentrations in 1958, so there are direct measurements showing the exponential rise of this gas (*1*) (Figure 1). Because CO_2 is a greenhouse gas, it is expected to impact on our future climate. By studying ancient climate, we can see the coupling between atmospheric CO_2 concentrations and major climate fluctuations. A remarkable record has been produced by scientists who extract bubbles of air trapped in Antarctic ice cores to produce a record that may go back as far as 1 million years. Figure 2 shows data to 350 thousand years before present. The cyclic changes in atmospheric CO_2 reflect glacial-interglacial cycles, with high CO_2 during warmer interglacial times, and low CO_2 during cold glacial times. Boron isotopes may allow us to extend the atmospheric CO_2 record back beyond the extent of ice core records. While not a direct measurement of atmospheric CO_2 concentration, it will give us the magnitude and frequency of past, pre-anthropogenic changes that can be related to climatic conditions determined by other proxies.

Understanding of Earth history and geological materials has been greatly facilitated by the discovery of fundamental chemical principles. It has been more than 200 years since Dalton first theorized on the structure of the atom and more than 100 years since Mendelev's 1870 version of the periodic table had reached the form close to that we use today. But it was Rutherford's 1911 "planetary" model and Neil Bohr's 1913 electrostatic model of the atom that allowed geologists to become geochemists and rigorously apply these new ideas to the study of the earth.

Geologic applications began with investigations of compounds that make up Earth materials. These are the minerals and rocks whose formation follows the basic chemical principles set forth by those pioneering scientists. V. M. Goldschmidt was a chemist who used the contemporary knowledge of these basic principles towards understanding crystal structures. His pioneering work to determine ionic radii and coordination number of elements in minerals led to many of the basic principles we use today in geochemistry, particularly in understanding the partitioning of elements in the Earth (see review in (*2*)). Goldschmidt is often called the "father of geochemistry" and is honored by the annual international meeting of the Geochemical Society, the most respected meeting for geochemists, the V.M. Goldschmidt Conference.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

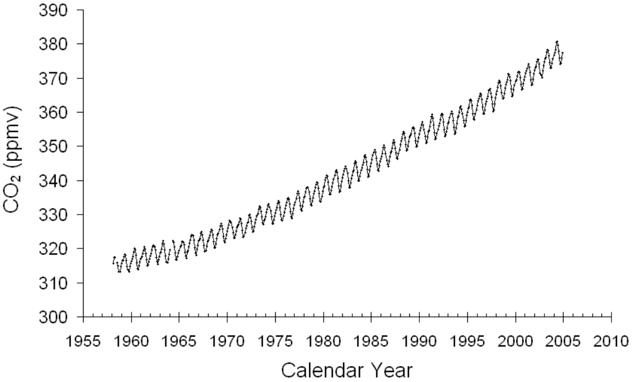


Figure 1. Atmospheric CO_2 concentrations measured on air samples from Mauna Loa. The dramatic increase since measurements began in 1958 are due to anthropogenic CO_2 additions predominantly from fossil fuel burning. Also captured in this record are seasonal cycles resulting from changes in northern hemisphere biomass (1, 3).

It is remarkable how quickly the field of isotope geochemistry evolved, considering the neutron was not identified until 1932! (*4*). Henri Becquerel was a physicist and mineralogist who identified products of radioactive decay in 1896, although this was not identified as radioactivity until Marie and Pierre Curie's work in Becquerel's lab. Urey's 1947 work on the thermodynamic basis for isotope exchange set the stage for the rapidly advancing field of stable isotope geochemistry. Calculation of the temperature dependence of oxygen isotope fractionation (*5, 6*) preceded the ability to measure isotopes precisely. Progressing from theory to the widespread application of these new concepts could not have happened without the development of analytical techniques to make the extremely precise measurements necessary to determine the small range in isotope compositions found in nature. While physicists led the way in developing early isotope instrumentation, Alfred Neir, who was also a geochemist, developed the isotope ratio mass spectrometer we know today (*7*).

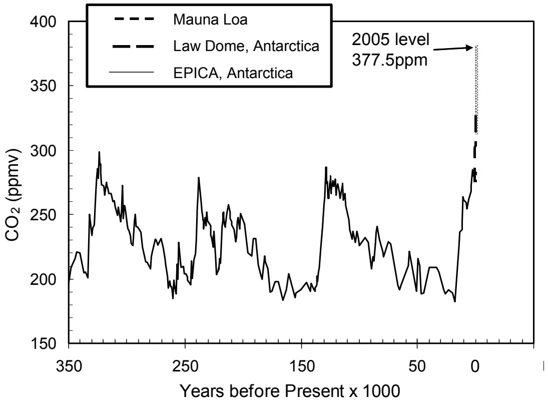


Figure 2. Atmospheric CO_2 concentrations over the last 350,000 years compiled from the Mauna Loa record (1957-2005) (1, 3), Law Dome ice core record (1010-1975) (8), Vostok ice core record (9).

Geochemists deserve much of the credit for advances in analytical techniques that pushed the limits of small sample size and improved analytical precision for Earth materials.

Applications of isotopes in geology fall into two broad categories: radiogenic isotope studies, where an unstable parent isotope decays to a daughter product, thus giving “age” information when analyzed in rocks and minerals, and stable isotope studies that generally rely on fractionation due to mass difference, and thus have been traditionally restricted to light isotope systems (typically elements with an atomic number <20). Here I will concentrate on mass dependent stable isotope systems.

The boron isotope system as a geochemical tool is still in the development phase. Presented here is a review of the basic chemical concepts behind the boron isotope paleo-*p*H proxy. A proxy in geochemistry is an indirect measurement of a parameter that is impossible to measure directly. Boron isotope compositions in calcium carbonate are predominantly controlled by the *p*H of the parent fluid, and therefore can be measured in ancient calcium carbonate-producing organisms (corals, shells of bivalves, calcifying micro-organisms, *etc.*) and used as a paleo-*p*H proxy. Any proxy needs to be thoroughly studied and tested to ensure other controls are not present, or can be corrected for. While presented at a very basic level, the fundamental chemical principles used to develop the boron isotope paleo-*p*H proxy present a starting point for a more in-depth study of these principles in a “real world” framework.

Following the fundamental studies on oxygen isotope systematics and the development of oxygen isotope paleo-thermometry, other geochemical isotope proxies have been proposed and developed at a fervent pace (see review in (*10*)). Boron isotopes were not investigated until late in the 1980’s, primarily due to analytical difficulties with this isotope system (described below). In order to develop this proxy, and in order to understand how and why this proxy works, knowledge of the following chemical principles is required:

1. basic chemistry including atomic theory and chemical reactions (dissolution-precipitation, bonding, trace elements, acid-base reactions);
2. aqueous speciation;
3. crystal chemistry, including trace element substitution in the crystal lattice;
4. isotope exchange reactions and mechanisms of isotope fractionation;
5. analytical chemistry, including ion exchange reactions and mass spectrometry (which includes ionization methods, in addition to the physics behind mass spectrometric analysis).

The application of these basic chemical principles to geologic and environmental problems will be discussed in reference to the boron isotope paleo-*p*H proxy.

Boron Isotope Systematics

Boron in seawater has a concentration of about 4.5ppm. It is supplied to the ocean by continental weathering and hydrothermal inputs, and has a residence

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

time of 14 to 20 Myr (*11-14*). The residence time of an element in the ocean is calculated by the total inventory of the element dissolved in the ocean divided by the total flux in (or out) of the ocean. Very soluble elements like Na and Cl have very long residence times (millions of years), and thus are concentrated in seawater. Further, they are well-mixed in seawater as the mixing time of the oceans is on thousand year time scales. Because of the long residence time of boron, ocean water has the same composition regardless of where it is collected around the world. Further, the boron isotopic composition is stable in the oceans over million-year time scales.

Boron is a Group IIIA element (+3 ion), but rarely occurs as a trivalent ion in nature. Its propensity to bind with oxygen results in primarily oxide and hydroxide compounds. Boron occurs as boric acid ($B(OH)_3$) in low *p*H solutions trigonally coordinated with the $(OH)^-$ radical. In high *p*H solutions, the borate ion is formed ($B(OH)_4^-$) in tetrahedral coordination. At intermediate *p*H values, both species are present, so consequently their relative proportions change with *p*H (Figure 3a). The fundamentally important observation relevant to the use of boron isotopes as a *p*H proxy is that these two aqueous species have an isotopic offset of about 20‰ (parts per thousand). Because the relative proportions of the aqueous species change with *p*H, so must the isotopic composition of those species, as the value of the reservoir remains constant (Figure 3b).

Why Does Boron Occur in the $CaCO_3$ Minerals?

Boron occurs as a trace element in $CaCO_3$, which forms two different crystal structures, the rhombohedral calcite structure and the orthorhombic aragonite structure. Both forms occur in nature, with biogenically-produced forms common in the oceans (*ie.* aragonitic coral skeletons, calcitic algae, aragonitic and/or calcitic bivalves, *etc.*). Because of the difference in crystal structure, trace element uptake in the two minerals is very different. For instance, Mg is readily incorporated into the calcite structure (often in the percent range in marine calcite), while it is relatively excluded from the aragonite structure, and Sr concentrations are low in calcite relative to aragonite. These differences stem from charge and size of the trace ions relative to the ions they are replacing, and follow basic chemical principles (*19-21*). This metal substitution in the Ca^{2+} site can be described by the distribution coefficient (see review in (*22*)). The distribution coefficient (K_D) for trace elements in carbonates is defined by:

$$K_D = \frac{[Me/Ca]_{solid}}{[Me/Ca]_{fluid}}$$

where Me is the metal that substitutes in the Ca site (molar concentrations). Boron is thought to substitute in the anion site in carbonates (*17, 18*), but the fact that calcite has significantly lower boron concentrations than aragonite is puzzling, as the opposite would be predicted based on ion size considerations. A model was proposed where only the charged $B(OH)_4^-$ species interacts with the crystal surface during crystal growth, but then must convert to a trigonal

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

coordination to be incorporated in calcite, resulting in an energy barrier to incorporation that explains the low uptake (18). No structural change is necessary for incorporation in aragonite. This interpretation stems from a study of boron coordination in carbonates, where it was shown that boron is primarily trigonal in calcite and dominantly tetrahedral in aragonite (23). The K_D equation was modified for boron substitution in the anion site (18):

$$K_{D^{B}_{CaCO_3}} = \frac{[B(OH)_3 / CO_3]_{solid}}{[B(OH)_4^- / HCO_3^-]_{fluid}}$$

This clearly cannot be the case for both aragonite and calcite, as the Sen *et al.* study showed (*23*). However, it is useful for considering boron uptake in carbonates. It is clear that the $B(OH)_4^-$ term controls the boron concentration in the carbonate, and so the concentration must be *p*H dependent. Higher *p*H results in higher $B(OH)_4^-$ in the fluid, resulting in higher boron concentrations in the crystal structure.

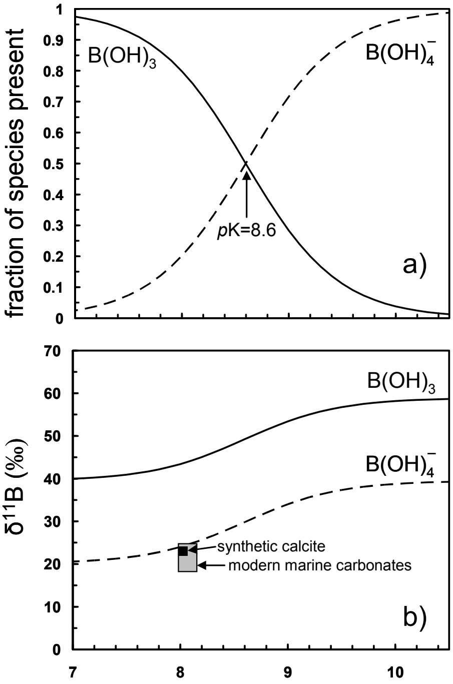


Figure 3. (a) Distribution of the two dominant aqueous species of boron in seawater calculated from pK_B value from (15). (b) The isotopic composition of the two dominant aqueous species of boron in seawater calculated using the theoretical fractionation factor of (16). The gray box represents the range of measured modern carbonates analyzed by (17), and include bivalves, corals, echinoderms, calcareous algae, and ooids, and the black box represents the boron isotopic composition of calcite, high magnesium calcite, and aragonite synthesized in the laboratory (18). The exact growth pH for the natural samples is not known, so values are plotted at typical surface seawater pH (8.2).

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

Why Are There Two Aqueous Species of Boron?

The aqueous speciation of boron is an acid-base reaction:

$$B(OH)_3 + (OH)^- \leftrightarrow B(OH)_4^-$$

Raising *p*H forces the reaction to the right following Le Chatelier's principle. The *p*H value where both aqueous species have equal concentrations is the *p*K (Figure 3a). This value has been determined by several investigators, but the accepted value is 8.60 at standard temperature and pressure (*15*) (see discussion in (*24*)). The *p*K is dependent on temperature, pressure and composition (salinity). Figure 4a shows the direction the *p*K value shifts using the natural extremes of these parameters, and Figure 4b the effect on the isotopic composition of the aqueous species. Under most natural conditions, factors besides *p*H are relatively insignificant, but must be taken into consideration when extremes in the environment are being studied, such as the surface ocean (relatively high T, low P) versus the deep ocean (relatively low T, high P).

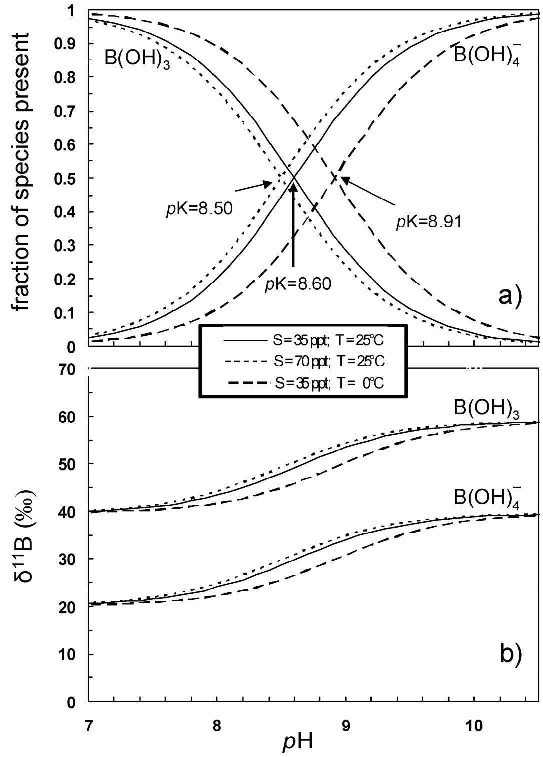


Figure 4. Change in pK values with temperature and salinity (a) and the effect on the isotopic composition of the aqueous boron species (b).

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

Why do the Two Predominant Aqueous Species Have Different Isotopic Compositions of Boron?

The fundamental chemical principle behind the isotopic offset between the two aqueous boron species relates primarily to vibrational and rotational energy differences between the two isotopes, such that the heavier isotope is preferentially incorporated into the trigonal species. The motivation for studying this fractionation was for the nuclear industry, as ^{10}B has a very high cross section for neutron capture, and so is used as a neutron flux absorber. Mechanisms for enriching ^{10}B from natural boron led to studies that showed it could be separated from ^{11}B on ion exchange columns. Theoretical approaches were used to calculate the fractionation factor α, which describes the isotopic offset between two phases,

$$\alpha = R_{B3}/R_{B4}$$

where R_{B4} is the isotopic composition of the tetrahedral $B(OH)_4^-$ species and R_{B3} is the isotopic composition of the trigonal $B(OH)_3$ species. The value for α that is generally used is 1.0194 (*16*), which is equivalent to an ~20‰ isotopic offset between these two species (Figure 3b). Because of uncertainties in the spectroscopic data used to determine the vibrational frequencies, large errors are present in that calculation. The theoretical approaches are elegantly reviewed and errors evaluated by (*25*), who concluded the Kakihana *et al.* value (*16*) is almost certainly too low, and called on scientists to experimentally determine α. Indeed, a recent experimental study to determine α concluded that it is significantly higher, 1.030 (*26*). However, empirical α values determined by measurement of the offset between seawater and calcium carbonate approximates the calculated fractionation factor of Kakihana *et al.*, (*17, 18, 27-29*); see Figure 5. These empirical studies were based on a major assumption, that only the $B(OH)_4^-$ aqueous species is incorporated in calcium carbonate (*17*). If this is correct, these studies give the value of the $B(OH)_4^-$ species for the particular *p*H at which the calcium carbonate formed, and thus can be used to calculate $\alpha_{empirical}$. Because of this consistency, the Kakihana *et al.* value (*16*) has been used for all of the boron isotope studies to date. An important point to make here is that, in geochemistry, where we are applying chemical principles to complex natural systems, deviations from the expected relationship are common. We use empirical relationships determined from laboratory and natural ground-truthing studies to establish the relationships, and to be confident that the proxy measurement is truly approximating the condition of interest.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

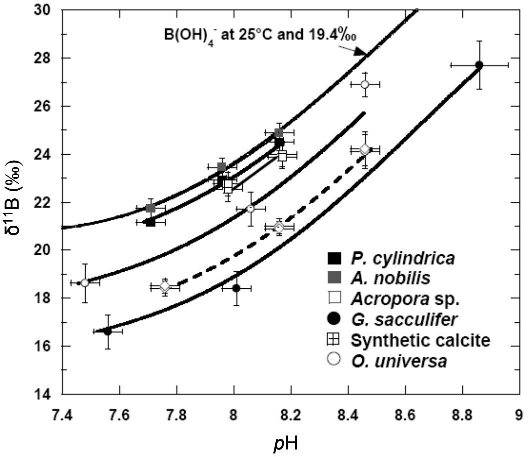


Figure 5. Ground-truthing studies showing the pH control on the boron isotopic composition of corals (P. cylindrica, A. nobilis, Acropora sp.), foraminifera (G. sacculifer, O. universa), and synthetic calcite compiled from (28-31). Note that the pH- $\delta^{11}B$ relationship of the samples define curves remarkably similar in shape to the theoretical $\delta^{11}B$ of the borate ion in seawater (arrow) determined by Kakihana et al. (16).

Ground-Truthing the Proxy

Testing theoretical geochemical relationships under controlled, but near-natural conditions, can provide convincing evidence that a proxy is approximating the parameter of interest. A brief review of both abiotic and biotic $CaCO_3$ experiments that have added to our knowledge of the boron isotope paleo-*p*H proxy is presented.

Establishing the *p*H-B Isotope Relationship

The first analyses of boron isotopes in carbonates were made to determine if boron in the solid recorded the isotopic composition of the seawater from which it precipitated (*17, 32*). These studies clearly showed an isotopic offset in the carbonates from that of seawater. In particular, the Hemming and Hanson study (*17*) showed a tight cluster of boron isotopic compositions ranging from about 20 to 25‰, significantly lighter than the ~+40‰ of seawater (Figure 3b). Note that isotopic values are in δ notation, which is the deviation of the isotope composition of a sample from that of a standard, which is defined as $\delta^{11}B_{standard} = 0$. The equation for calculating δ is:

$$\delta^{11}B = \frac{\left({}^{11}B/{}^{10}B\right)_{SAMPLE} - \left({}^{11}B/{}^{10}B\right)_{STANDARD}}{\left({}^{11}B/{}^{10}B\right)_{STANDARD}} \bullet 1000$$

Therefore, positive values are isotopically heavier than the standard, and negative values are isotopically lighter than the standard. The accepted analytical standard is a boric acid powder supplied by the National Institute of Standards and Technology (NIST SRM 951).

The explanation for this large fractionation was not immediately obvious. Temperature-controlled fractionation could not produce such a large offset. The idea that the fractionation could be controlled by a process related to the aqueous speciation stemmed from a previous study evaluating the large isotopic difference seen in terrestrial versus marine materials. Early studies found that terrestrially-derived borate deposits were isotopically light compared to marine-derived borate deposits (see (*33*)). An earlier study proposed a model that explains this offset (*34*). When boron enters the oceans from the neutral to acidic riverine sources (isotopically light, terrestrial boron), it encounters the relatively basic (*p*H~8.2) ocean water, where the boron is distributed between the neutrally charged $B(OH)_3$ species (~75% of the boron) and the negatively charged $B(OH)_4^-$ species (~25% of the boron). As shown in the relationships in Figures 3a and 3b, this process produces an isotopic offset between the two aqueous species. The charged aqueous species adsorbs to clays that are also entering the oceans with the riverine input. This effectively sequesters the isotopically light aqueous species, thus leaving seawater boron isotopically heavy (*34*). It was later shown that similar processes take place during hydrothermal alteration of ocean ridge basalts (*14*). This same model was used to suggest that during precipitation of calcium carbonate from the oceans, only the $B(OH)_4^-$ species is interacting with the calcite or aragonite crystal surface (*17*). This makes intuitive sense, as the uncharged $B(OH)_3$ species is unlikely to react with the charged crystal surface. If this model is correct, then the calcium carbonate should have the same boron isotopic composition as the borate ion in seawater, and indeed the cluster of compositions of modern marine carbonates strongly supports this model (Figure 3b). The range of values is easily explained by the fact that the survey studies did not have a record of the ambient *p*H for each of the samples. In order to rigorously test this model, the *p*H of the water had to be known. This led to a laboratory-based study, where calcite and aragonite were synthesized at a known and constant *p*H in a solution with similar ionic strength to seawater (*18*). In that study, both the calcite and aragonite crystals had isotopic compositions that clustered over a very small range (± 0.7 ‰), and very close to the $B(OH)_4^-$ curve (Figure 3b).

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

Empirical Relationships

This set the stage for further studies to determine the empirical relationship for seawater/calcium carbonate fractionation in natural samples. Several studies of the boron isotopic compositions of corals cultured under controlled conditions showed that the relationship found in the Hemming *et al.* study (*18*) held for corals (*28, 30*). However, studies of foraminifera, a calcifying single celled marine organism, showed species dependent and variable offsets from the $B(OH)_4^-$ curve (*27, 29*). These studies collected foraminifera in their juvenile state taken from the surface ocean near Catalina Island, California. The juvenile specimens were taken to the laboratory where they calcified under controlled *p*H conditions attained by adjusting alkalinity in the growth chamber. While these studies verified the strong *p*H dependence on the isotopic composition of shell material, they typically fell significantly below the theoretical $B(OH)_4^-$ curve (Figure 5). This was attributed to "vital effects", a term commonly used in geochemistry to explain apparent deviations from equilibrium induced by biological processes (*ie.* photosynthesis). These arise from the fact that minimal energy states achieved by equilibrium reactions are not always consistent with the processes organisms require for their energy needs. Further study will be necessary to fully understand these results. However, it is remarkable how consistently the shape of the empirical curves mimic that of the Kakihana *et al.* $B(OH)_4^-$ curve. This has led to the conclusion by those who established the empirical data, as well as theoretical studies of this system (*25*), that once an empirical curve is established for a particular species, the boron *p*H proxy can be used with confidence if samples are carefully chosen. With this knowledge, an equation can be used to calculate the *p*H from the boron isotope measurement and the known species-specific offset:

$$pH = pK_B - \log - \left[\frac{\delta^{11}B_{seawater} - \delta^{11}B_{sample} - a}{\delta^{11}B_{seawater} - \alpha * \left(\delta^{11}B_{sample} + a\right) - \varepsilon} \right]$$

where $\delta^{11}B_{seawater} = +39.5$, ε is the fractionation between the aqueous boron species (19.2‰) and a is the empirically derived species specific offset between $\delta^{11}B$ of dissolved borate and that of the foraminifera shell.

To fully test the boron isotope paleo-*p*H proxy, controls other than *p*H that might affect the boron isotope composition must either be measured and corrected, or minimized in some way. Temperature was not expected to be a significant control under normal surface ocean temperature ranges based on theoretical predictions (*16*). We tested this by the same culture experiments described previously, and indeed found no control of temperature in coral samples outside the uncertainty of the analyses (*28*) over a small temperature range (25-28ºC). Further, measurement of the foraminifera species *G. sacculifer* over the range ~25-28ºC again found no temperature control (*27*). Other possible complications were also investigated. Foraminifera do not reside near the ocean surface during their entire lifecycle, and in fact some individuals reside deeper in the water column. Those individuals will record *p*H at their habitat depth (*p*H typically decreases with depth in the water column). It was found that individuals residing deeper in the water column produced smaller shells. This effect was evaluated for boron isotopes, and indeed smaller individuals had isotopically lighter boron isotope compositions consistent with the deeper, lower *p*H waters in which they resided (*35*). In that same study, dissolution effects were investigated. Dissolution can occur in the deep ocean, as decomposition of organic matter and the high pCO_2 attained due to the high pressure in the deep ocean makes bottom waters more acidic. Thus, when the shells fall to the ocean floor they may partially dissolve. It was found that partial dissolution also lowered the boron isotope composition, making these shells not useful for the boron isotope *p*H proxy (*35*).

While more study is necessary to further refine this proxy, it is clear that the major control on the boron isotope composition of carbonates is the *p*H of the

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

water in which they grew, and that careful sample selection (large, heavy foraminifera) will minimize complications that may give erroneous results.

Analytical Aspects

A brief discussion of the necessary steps from sample collection to data collection affords an opportunity to present basic chemical principles applied to the boron isotope paleo-*p*H proxy. While extraction of boron from samples is the first step in the procedure, it is useful to discuss the mass spectrometry first, as the extraction method is different for the various mass spectrometric methods used.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

Mass Spectrometry

While the ratio of the stable boron isotopes is such that it would seem to be relatively easy to measure compared to other stable isotope systems ($^{11}B/^{10}B \approx 4$, compared to $^{18}O/^{16}O \approx 0.002005$ and $^{13}C/^{12}C \approx 0.011122$), the chemical properties of boron required modification of standard techniques for isotope measurement. In particular, mass fractionation induced during thermal ionization in the mass spectrometer produces large analytical uncertainties in the measurements. This was overcome to some degree by producing heavier ionic compounds (*ie.* $Na_2BO_2^+$, $Cs_2BO_2^+$ (*33, 36-41*)), or by carefully replicating analytical conditions for standards and samples measuring the BO_2^- ion (*17, 42-44*). Each of these approaches has advantages and disadvantages. The positive ions produced by the heavier compounds result in significantly lower ion yields compared to the negatively charged ions. Typically micrograms of boron are necessary for the positive ion techniques, while only nanograms of boron are necessary for the negative ion technique. Comparison of these procedures can be found in (*45*) and (*42*).

Recent advances in multi-collector inductively-coupled plasma mass spectrometry (MC-ICPMS) techniques show promise, as this method allows bracketing of samples by standards, thus providing a means for near-simultaneous fractionation correction. Problems with this technique include memory effects (boron from previously analyzed samples or standards remobilized during subsequent analyses), and extreme sensitivity to the solution matrix, requiring samples and standards to have identical matrices.

Sample Preparation

Boron can be extracted from natural or synthetic calcium carbonate (or other geologic materials) before analysis by mass spectrometry. The extraction method depends on the analytical technique to be used. For any method, natural samples (typically foraminifera, a single-celled marine invertebrate whose shells accumulate in ocean sediments) must be cleaned of organic matter. This is typically done by ultrasonification of the crushed samples in the presence of a strong oxidizing agent such as H_2O_2 (hydrogen peroxide) or NaOCl (sodium hypochlorite, or common household bleach). For the negative ion method, efficient ionization actually depends on the presence of a salt on the filament which acts as an electron donor, so sample preparation is quite simple. Dissolving the calcium carbonate in HCl provides the necessary salts, so this solution can be loaded directly on a metal filament (typically Re) for mass spectrometric analysis. Preparation for the positive ion and MC-ICPMS methods, however, require quantitative isolation of boron from the dissolved sample. Quantitative separation and recovery of the boron is essential, as fractionation of the isotopes occurs if <100% of the boron is recovered, and inefficient separation of other ions (*i.e.* Ca) causes matrix-induced effects during mass spectrometry, such as changing the mass spectrometer-induced fractionation, or interfering with ionization of the boron thus resulting in low ion yields.

Most separation techniques use ion exchange columns to separate the boron (*17, 42, 47*). The ion exchange methods all work by the same fundamental chemical principles, although details of the chemistry differ depending on the type of ion-exchange material being used. Commonly, a boron-specific ion-exchange material is preferred (Amberlite© IRA-743 anion exchange resin), due to its high affinity for boron, and low affinity for most other unwanted ions. This boron-specific resin has exchange sites that allow $B(OH)_4^-$ to exchange for the OH^- radical, thus sequestering the boron on the resin bed (Figure 6).
The procedure for the small (1-5mg) calcium carbonate samples we analyze is as follows:

1. load ~20μL of the boron-specific resin into a column made from a 1ml polypropylene pipette tip with a porous polypropylene frit pressed into the tip to support the ion-exchange beads;
2. clean with 20 column volumes of 0.2N HCl to remove any "environmental" boron;
3. rinse with 20 column volumes deionized water;
4. condition with 5 column volumes base (1M NH_3) to provide the tertiary OH^- at the resin bead exchange site, then rinse with 5 column volumes deionized water;
5. load sample on the column (at neutral *p*H) and wash with 5 column volumes dionized water;
6. elute the boron with 10 column volumes low *p*H solution (0.1 N HCl or other acid).

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

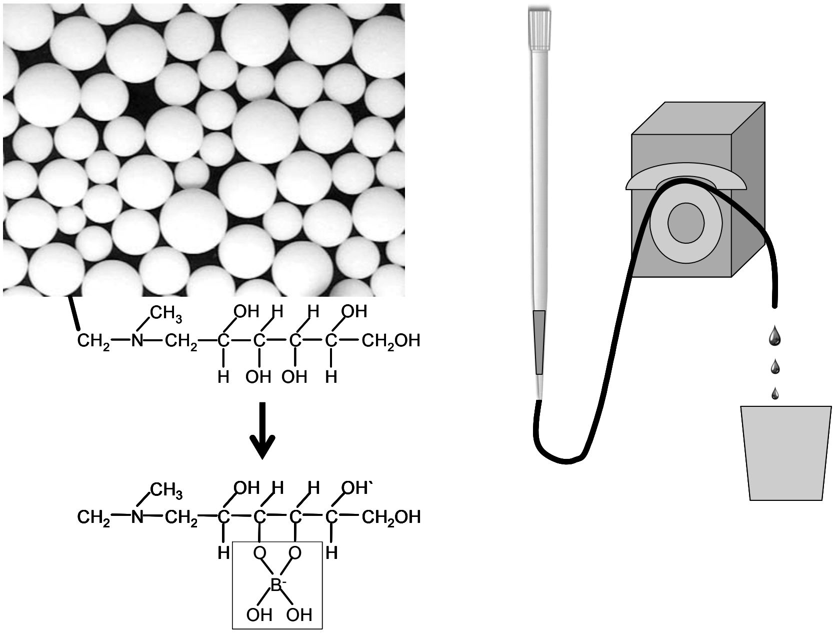


Figure 6. Schematic of the ion-exchange apparatus for isolating boron from geologic samples. A peristaltic pump draws the dissolved sample and reagents through an ion exchange column fabricated from a 1 mL polypropylene pipette tip packed with boron specific resin (IRA-743). The photograph of the actual beads is from the manufacturers technical report (Rohm and Haas). The mechanism of exchange of borate on the tertiary amine groups on the resin surface is also shown (46).

This procedure produces a pure boric acid solution that is ready for analysis by any of the mass spectrometric techniques. Due to the high affinity for the B-specific resin, it is difficult to release the boron from the ion-exchange column, thus requiring the large volume of acid (step 6). As a result, concentration by evaporation may be necessary, particularly for the positive ion technique. For negative ion analysis, since salts were removed, a salt must be added back in. This is typically done by adding a boron-free seawater solution (*42*), prepared by putting natural seawater through the exact same extraction method as described above, except of course collecting the sample from step 5 rather than step 6. The sample boron and the boron-free seawater can be combined directly on the ionization filament. The negative ion method has analytical uncertainties in the 0.3-0.5‰ range if samples are analyzed multiple times (typically three replicates). Because of the relatively low masses measured ($^{11}BO_2^-/^{10}BO_2^-$, masses 43 and 42), instrumental fractionation is large. Since there is no established method to correct for this, we rely on analyses of standards to determine the fractionation from the true value. A strict analytical protocol has been established to ensure reproducibility of the standard (and thus samples), which includes a consistent ionization filament temperature (980°C), limits on in-run changes in fractionation, and a minimum of three replicate analyses that meet those criteria (see discussion in (*31*)).

For the positive ion technique, the accepted method today is to analyze $Cs_2BO_2^+$ (masses 208 and 209). Extracted boric acid from the column chemistry procedure is mixed with a cesium hydroxide solution on the ionization filament (along with a graphite slurry) prior to loading in the mass spectrometer. Reduced instrument-induced mass fractionation during the analysis due to the very high mass number has typically resulted in analytical uncertainties in the 0.2-0.5‰ range for a single analysis.

Applications to the Ancient Ocean and Atmosphere

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

Establishing the clear relationship between the boron isotope composition of marine carbonates and the *p*H of seawater set the stage for determining *p*H of ancient oceans. While a review of these applications has been presented previously (see (*44*)), a brief discussion here will present the final (and most important) step in the proxy development story.

Determining ocean *p*H in the past is a difficult but extremely important measurement which allows estimations of ancient atmospheric CO_2. Understanding natural atmospheric CO_2 is necessary for understanding the consequences of the dramatic anthropogenic atmospheric CO_2 increase we are now experiencing (Figure 1). We now have a record of atmospheric CO_2 going back about 750 thousand years (*9*). This record was produced by extracting bubbles of trapped air from the Vostok ice core in Antarctica, then analyzing their CO_2 concentrations. In addition, other proxy measurements of the same ice core allowed determination of air temperature fluctuations over the poles. The Antarctic ice cores provide a remarkable record that demonstrate the tight coupling between air temperature and atmospheric CO_2 concentrations, aiding our understanding of glacial-interglacial cycles (Figure 7). This type of work is essential to determine natural fluctuations in atmospheric CO_2 and how those fluctuations relate to global climate conditions. There is a strong correlation between the glacial cycles and the cyclic change in the amount and distribution of solar radiation reaching the earth. The “Milankovitch” cycles allow estimations of solar radiation influx as it changes with eccentricity (distance), obliquity (tilt of Earth’s axis), and precession (wobble of the Earth around its axis), with cycle lengths of 100,000, 41,000, and 23,000 years, respectively. The phasing of these three parameters has been used to calculate insolation over the past 1 million years (*48*). When total radiation influx is lowest, the Earth enters a glacial or “icehouse” stage, and when it is highest, it is in an interglacial or “greenhouse” stage. However, it has also been shown that these changes in solar energy alone are not capable of forcing the Earth in and out of glacial periods. It must be that it affects (or “triggers”) some other climate forcing mechanism, which amplifies the solar radiation cycles. Atmospheric CO_2 is thought to be that amplifier, as there are mechanisms for redistributing carbon from various reservoirs on the earth (atmosphere, oceans, rocks, plants). Climate scientists believe that understanding this transfer of carbon is key to understanding the major climate changes we see over the last several million years.

Figure 7 clearly illustrates the tight coupling between atmospheric CO_2 concentrations and proxies for temperature and ice volume. The isotopic composition of oxygen ($\delta^{18}O$) in marine calcifying organisms (in this case the single-celled foraminifera *G. Ruber*), is a function of two variables: the water temperature and the extent of continental glaciation. Evaporation of seawater results in an isotopic fractionation between the liquid water and water vapor. Since water vapor is significantly lighter, accumulation of precipitation as ice during glacial times leaves seawater isotopically heavy, which is recorded in the skeletons of marine organisms. Because there is another proxy for temperature that can be measured in foraminifera, it is possible to extract an "ice volume" record from the ocean sediments. Mg/Ca in shells changes with temperature, with less discrimination against incorporation of the trace element Mg as temperature increases. Figure 7c is the seawater oxygen isotope record with the temperature component removed, so it is essentially a record of changes in the extent of glaciation. Comparing that record to the Vostok atmospheric CO_2 record, it is clear there is a coupling, with low CO_2 correlated with high ice volume. While the easy explanation is the greenhouse/icehouse scenario (high CO_2 traps heat and causes "greenhouse" conditions, low CO_2 cools the earth causing "icehouse" glacial conditions), it may not be that simple. Cause and effect is often difficult to determine in ancient records, particularly because of imprecision in dating the sediments or ice from which the records come.

To determine atmospheric CO_2 changes prior to about 1 million years before present (the probable limit of ice core data), a proxy is necessary. The first application of the boron isotope proxy to deep time was that of (*52*). This study analyzed boron isotopes in foraminifera (a single-celled marine calcifier) taken from ocean cores with sediments dating back to 21 Myr. While this study was an important first step in applying this proxy, later studies revealed that there is a species dependent offset (see discussion on ground-truthing above and Figure 5) that needs to be determined for each type of foraminifer analyzed (the Spivack *et al.* study used mixed foraminifera species (52)). Sanyal et al. analyzed a single species of surface-dwelling foraminifera, measuring a glacial-interglacial offset of ~0.3 pH units, which was used to calculate the atmospheric pCO2 in equilibrium with that pH value (*53*). The result was a pCO_2 value of about 200 ppm, consistent with the known glacial pCO_2 measured from air trapped in ice cores. Subsequent studies used *p*H values interpreted from boron isotope measurements to study ancient upwelling, where low *p*H deep water comes to the surface (*54, 55*), and to determine *p*H depth profiles over the past 16 Myr (*56*). For aminifera samples were measured going back 43 and 60 Myr (*57, 58*). However, since this spans a time much greater than the residence time of boron in the oceans (~16 Myr), the ocean boron isotope composition cannot be assumed to be constant (see discussion in (*12, 13*)). Relative *p*H values are probably reliable, but absolute *p*H values may not be possible until we have determined the secular variation of the ocean boron isotope composition.

Hönisch *et al.* applied all of the knowledge obtained from the ground-truthing studies, and limited their analyses to the past two glacial cycles (Figure 7) so that the data could be compared directly to ice core records (*50*). The

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

results of this study showed an agreement between atmospheric pCO_2 calculated from the boron isotope pH proxy and the ice core measurements to within ± 30 ppmV. The next step is to take the proxy further back than the ice core records, and a study in progress will determine atmospheric pCO_2 to 1.5Ma.

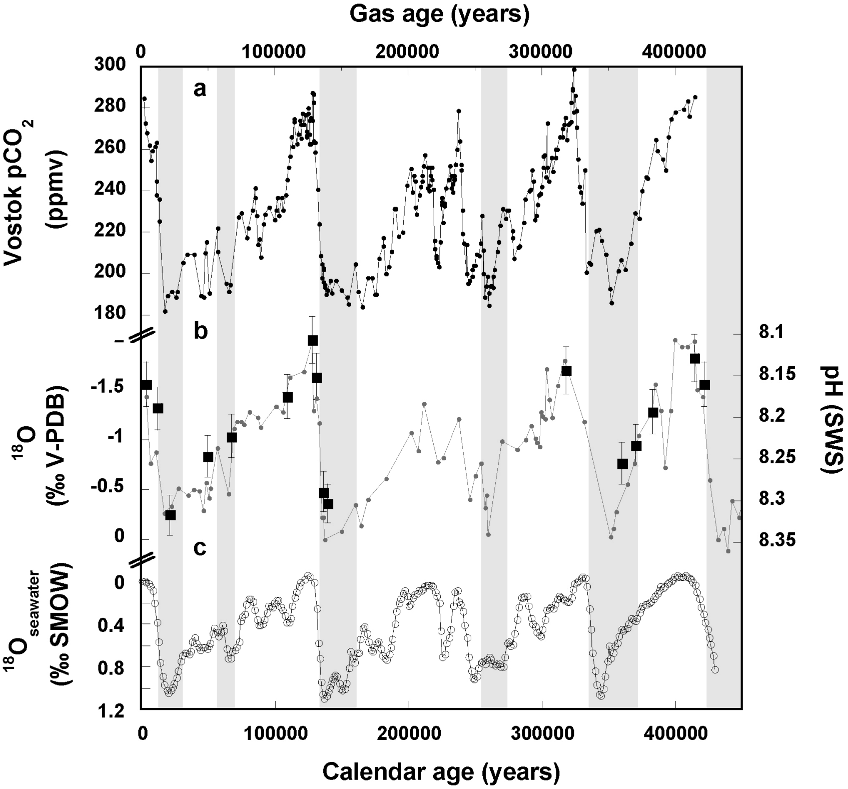


Figure 7. (a) Vostok ice core CO_2 record reported as ppmv (part per million by volumbe) (9). (b) Ocean sediment core (G. ruber, a species of foraminifera) oxygen isotope record reported as the deviation in parts per thousand from the marine carbonate standard V-PDB (both sea-surface temperature and changes in the bulk seawater oxygen isotope composition affect this record) (49), and boron isotope based pH values reported relative to the seawater pH scale (large squares) (50). (c). Continental ice volume signal from $\delta^{18}O$ seawater reported as the deviation in parts per thousand from the seawater standard SMOW. This record is reconstructed by removing the temperature signal from the sediment core record, and assuming changes in the seawater oxygen isotope composition are due to the storage of light oxygen in continental ice sheets (51). (a) and (b) are on the Vostok gas age scale, and (c) is calendar age. Gray bars indicate glacial intervals. Figure reprinted from (50) with permission from Elsevier.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

References

1. Keeling, C. D.; Piper, S. C.; Bacastow, R. B.; Wahlen, M.; Whorf, T. P.; Heimann, M.; Mayer, M. G. In *A History of Atmospheric CO_2 and its effects on Plants, Animals, and Ecosystems;* Ehleringer, J. R.; Cerling, T. E.; Dearing, M. D., Eds.; Ecological Studies; Springer Verlag: New York, NY, 2005; Vol. 177, pp 83-113 .
2. Mason, B. *Victor Moritz Goldschmidt: Father of Modern Geochemistry;* Special Publication No. 4; The Geochemical Society: San Antonio, TX, 1992.
3. Keeling, C. D. Whorf, T. P. In *Trends: A compendium of data on global change;* U.S. Department of Energy: Oak Ridge, TN, 2005.
4. Chadwick, J. The existence of a neutron. *Proc. Roy. Soc.* **1932,** *136,* 692-708.
5. Urey, H. C. The thermodynamic properties of isotopic substances. *J. Chem. Soc.* **1947,** 562-581.
6. Bigeleisen, J.; Mayer, M. G. Calculation of equilibrium constants for isotopic exchange reactions. *J. Chem. Phys.* **1947,** *15,* 261-267.
7. Neir, A. A mass spectrometer for routine isotope abundance measurements. *Rev. Sci. Instrum.* **1940,** *11,* 212-216.
8. Etheridge, D. M.; Steele, L. P.; Langenfelds, R. L.; Francey, F. J.; Barnola, J.-M.; Morgan, V. I. Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn. *J. Geophys. Res.* **1996,** *101,* 4115-4128.
9. Petit, J. R.; Jouzel, J.; Raynaud, D.; Barkov, J. M.; Barnola, J. M.; Basile, I.; Bender, M.; Chappellaz, J.; Davis, M.; Delaygue, G.; Delmotte, M.; Kotlyakov, V. M.; Legrand, M.; Lipenkov, V. Y.; Lorius, C.; Pepin, L.; Ritz, C.; Saltzman, E.; Stievenard, M. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **1999,** *399,* 429-436.
10. Henderson, G. M. New oceanic proxies for paleoclimate. *Earth Planet. Sci. Lett.* **2002,** *203,* 1-13.
11. Taylor, S. R.; McLennan, S. M. *The continental crust: Its composition and evolution;* Blackwell Scientific Publications, Palo Alto, CA, 1985.
12. Lemarchand, D.; Gaillardet, J.; Lewin, E.; Allegre, C. J. Boron isotope systematics in large rivers: implications for the marine boron budget and paleo-*p*H reconstruction over the Cenozoic. *Chem. Geol.* **2002,** *190,* 123-140.
13. Lemarchand, D.; Gaillardet, J.; Lewin, É.; Allègre, C. J. The influence of rivers on marine boron isotopes and implications for reconstructing past ocean *p*H. *Nature* **2000,** *408,* 951-954.
14. Spivack, A. J.; Palmer, M. R.; Edmond, J. M. The sedimentary cycle of the boron isotopes. *Geochim. Cosmochim. Acta* **1987,** *51,* 1939-1949.
15. Dickson, A. G. Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K. *Deep-Sea Res.* **1990,** *37,* 755-766.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

16. Kakihana, H.; Kotaka, M.; Satoh, S.; Nomura, M.; Okamoto, M. Fundamental studies on the ion-exchange of boron isotopes. *Bull. Chem. Soc. Jpn.* **1977,** *50,* 158-163.
17. Hemming, N. G.; Hanson, G. N. Boron isotopic composition and concentration in modern marine carbonates. *Geochim. Cosmochim. Acta* **1992,** *56,* 537-543.
18. Hemming, N. G.; Reeder, R. J.; Hanson, G. N. Mineral-fluid partitioning and isotopic fractionation of boron in synthetic calcium carbonate. *Geochim. Cosmochim. Acta* **1995,** *59,* 371-379.
19. Goldschmidt, V. M.; Barth, T.; Lunde, G.; Zachariasen, W. Geochemische Verteilungsgesetae, VII: Die Gesetze der Krystallochernie. *Nat. Kl* **1926.**
20. Shannon, R. D.; Prewitt, C. T. Effective ionic radii and crystal chemistry. *Journal of Inorganic and Nuclear Chemistry* **1970a,** *32,* 1427-1441.
21. Shannon, R. D.; Prewitt, C. T. Revised values of effective ionic radii. *Acta Crystallogr., Sect. B: Struct. Sci.* **1970b,** *B26,* 1046-1048.
22. Morse, J. W.; Bender, M. L. Partition coefficients in calcite: Examination of factors influencing the validity of experimental results and their application to natural systems. *Chem. Geol.* **1990,** *82,* 265-277.
23. Sen, S.; Stebbins, J. F.; Hemming, N. G.; Ghosh, B. Coordination environments of B impurities in calcite and aragonite polymorphs: A ^{11}B MAS NMR study. *Am. Mineral.* **1994,** *79,* 819-825.
24. Zeebe, R. E.; Wolf-Gladrow, D. A. *CO_2 in seawater: Equilibrium, kinetics, isotopes;* Elsevier Oceanography Series, 65; Elsevier: Amsterdam, 2001.
25. Zeebe, R. E. Stable boron isotope fractionation between dissolved $B(OH)_3$ and $B(OH)_4^-$. *Geochim. Cosmochim. Acta* **2005,** *69,* 2753-2766.
26. Byrne, R. H.; Yao, W.; Klochko, K.; Tossell, J. A.; Kaufman, A. J. Experimental evaluation of the isotopic exchange equilibrium $^{10}B(OH)_3+^{11}B(OH)_4^-=^{11}B(OH)_3+^{10}B(OH)_4^-$ in aqueous solution. *Deep Sea Research Part I: Oceanographic Research Papers* **2006,** *53,* 684-688.
27. Hönisch, B.; Hemming, N. G. Ground-truthing the boron isotope paleo-*p*H proxy in planktonic foraminifera shells: Partial dissolution and shell size effects. *Paleoceanography* **2004,** *19,* PA4010.
28. Reynaud, S.; Hemming, N. G.; Juillet-Leclerc; A. Gattuso, J.-P. Effect of pCO_2 and temperature on the boron isotopic composition of a zooxanthellate coral: *Acropora* sp. *Coral Reefs* **2004,** *23,* 539-546.
29. Sanyal, A.; Hemming, N. G.; Broecker, W. S.; Lea, D. W.; Spero, H. J.; Hanson, G. N. Oceanic *p*H control on the boron isotopic composition of foraminifera: Evidence from culture experiments. *Paleoceanography* **1996,** *11,* 513-517.
30. Hönisch, B.; Hemming, N. G.; Grottoli, A. G.; Amat, A.; Hanson, G. N.; Bijma, J. Assessing scleractinian corals as recorders for paleo-*p*H: Empirical calibration and vital effects. *Geochim. Cosmochim. Acta* **2004,** *68,* 3675-3685.
31. Hönisch, B.; Hemming, N. G.; Loose, B. Comment on "A critical evaluation of the boron isotope-pH proxy: The accuracy of ancient ocean pH estimates" by M. Pagani, D. Lemarchand, A. Spivack and J. Gaillardet. *Geochim. Cosmochim. Acta* **2007,** *71,* 1636-1641.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

32. Vengosh, A.; Kolodny, Y.; Starinsky, A.; Chivas, A. R.; McCulloch, M. T. Coprecipitation and isotopic fractionation of boron in modern biogenic carbonates. *Geochim. Cosmochim. Acta* **1991,** *55,* 2901-2910.
33. Swihart, G. H.; Moore, P. B.; Callis, E. L. Boron isotopic composition of marine and nonmarine evaporite borates. *Geochim. Cosmochim. Acta* **1986,** *50,* 1297-1301.
34. Schwarcz, H. P.; Agyei, E. K.; McMullen, C. C. Boron isotopic fractionation during clay adsorption from sea-water. *Earth Planet. Sci. Lett.* **1969,** *6,* 1-5.
35. Hönisch, B.; Bijma, J.; Russell, A. D.; Spero, H. J.; Palmer, M. R.; Zeebe, R. E.; Eisenhauer, A. The influence of symbiont photosynthesis on the boron isotopic composition of foraminifera shells. *Mar. Micropaleontol.* **2003,** *49,* 87-96.
36. Leeman, W. P.; Vocke, R. D.; Beary, E. S.; Paulsen, P. J. Precise boron isotopic analysis of aqueous samples: Ion exchange extraction and mass spectrometry. *Geochim. Cosmochim. Acta* **1991,** *55,* 3901-3907.
37. Nakamura, E.; Ishikawa, T.; Birck, J.-L.; Allègre, C. J. Precise boron isotopic analysis of natural rock samples using a boron-mannitol complex. *Chem. Geol. (Isotope Geosci. Sect.)* **1992,** *94,* 193-204.
38. Oi, T.; Nomura, M.; Musashi, M.; Ossaka, T.; Okamoto, M.; Kakihana, H. Boron isotopic compositions of some boron minerals. *Geochim. Cosmochim. Acta* **1989,** *53,* 3189-3195.
39. Ramakumar, K. L.; Parab, A. R.; Khodade, P. S.; Almaula, A. I.; Chitambar, S. A.; Jain, H. C. Determination of isotopic composition of boron. *J. Radioanal. Nucl. Chem.* **1985,** *94,* 53-61.
40. Spivack, A. J.; Edmond, J. M. Determination of boron isotope ratios by thermal ionization mass spectrometry of the dicesium metaborate cation. *Anal. Chem.* **1986,** *58,* 31-35.
41. Xiao, Y.-K.; Beary, E. S.; Fassett, J. D. An improved method for the high-precision isotopic measurement of boron by thermal ionization mass spectrometry. *Int. J. Mass spectrom.* **1988,** *85,* 203-213.
42. Hemming, N. G.; Hanson, G. N. A procedure for the isotopic analysis of boron by negative thermal ionization mass spectrometry. *Chem. Geol.* **1994,** *114,* 147-156.
43. Vengosh, A.; Chivas, A. R.; McCulloch, M. T. Direct determination of boron and chlorine isotopes in geological materials by negative thermal-ionization mass spectrometry. *Chem. Geol.* **1989,** *79,* 333-343.
44. Hemming, N. G.; Hönisch, B. In *Proxies in Late Cenozoic Paleoceanography;* Hillaire-Marcel, C.; De Vernal, M., Eds.; Developments in Marine Geology 1; Elsevier: Amsterdam, 2007, pp 717-734.
45. Swihart, G. H. In *Boron: Mineralogy, Petrology and Geochemistry;* Grew, E. S.; Anovitz, L. M., Eds.; Reviews in Mineralogy, Vol. 33; Mineralogical Society of America: Washington, D.C.1996, pp 845-862.
46. Baek, K. W.; Song, S. H.; Kang, S. H.; Rhee, Y. W.; Lee, S. S.; Lee, B. J.; Hudson, S.; Hwang, T. S. Adsorption kinetics of boron by anion exchange resin in packed column bed. *J. Ind. Eng. Chem.* **2007,** *13,* 452-456.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

47. Kiss, E. Ion-exchange separation and spectrophotometric determination of boron in geological materials. *Anal. Chim. Acta* **1988,** *211,* 243-256.
48. Berger, A.; Loutre, M. F. Pre-Quaternary Milankovitch frequencies. *Nature* **1989,** *342,* 133.
49. Bird, M. I.; Cali, J. A. A revised high-resolution oxygen-isotope chronology for ODP-668B: Implications for Quaternary biomass burning in Africa. *Global and Planetary Change* **2002,** *33,* 73-76.
50. Hönisch, B.; Hemming, N. G. Surface ocean *p*H response to variations in pCO_2 through two full glacial cycles. *Earth Planet. Sci. Lett.* **2005,** *236,* 305-314.
51. Waelbroeck, C.; Labeyrie, L.; Michel, E.; Duplessy, J. C.; McManus, J. F.; Lambeck, K.; Balbon, E.; Labracherie, M. Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **2002,** *21,* 295-305.
52. Spivack, A. J.; You, C.-F.; Smith, H. J. Foraminiferal boron isotope ratios as a proxy for surface ocean *p*H over the past 21 Myr. *Nature* **1993,** *363,* 149-151.
53. Sanyal, A.; Hemming, N. G.; Hanson, G. N.; Broecker, W. S. Evidence for a higher *p*H in the glacial ocean from boron isotopes in foraminifera. *Nature* **1995,** *373,* 234-236.
54. Palmer, M. R.; Pearson, P. N. A 23,000-year record of surface water *p*H and PCO_2 in the Western Equatorial Pacific Ocean. *Science* **2003,** *300,* 480-482.
55. Sanyal, A.; Bijma, J. A comparative study of northwest Africa and eastern equatorial Pacific upwelling zones as sources of CO_2 during glacial periods based on boron isotope paleo-*p*H estimation. *Paleoceanography* **1999,** *14,* 753-759.
56. Palmer, M. R.; Pearson, P. N.; and Cobb, S. J. Reconstructing past ocean *p*H-depth. profiles. *Science* **1998,** *282,* 1468-1471.
57. Pearson, P. N.; Palmer, M. R. Middle Eocene seawater *p*H and atmospheric carbon dioxide concentrations. *Science* **1999,** *284,* 1824-1826.
58. Pearson, P. N.; Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **2000,** *406,* 695-699.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch009

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

Chapter 10

The Evolution of Chemistry through Synthesis (and of Synthesis in Chemistry)

Thomas R. Hoye* and Elena Sizova

Department of Chemistry, University of Minnesota,
Minneapolis, MN 55455

Other than during the last, say, two centuries, the synthesis/production of new molecules has taken place nearly exclusively in the domain of biology. Human intervention now complements natural evolution as a means for creating novel substances. Synthetic chemists have used higher order skills to push the limits both of *what* we make as well as of *how* we construct molecular entities. In this chapter we offer some perspective on the evolution of synthetic organic chemistry. Enabling technologies and intellectual developments are presented, and the synthesis of natural products is used as a template to demonstrate the growth of the field.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Introduction

The human condition today has benefited from successes in organic synthesis. Our food, clothing, homes, and health have been significantly improved by the creation of new substances based on carbon compounds. Safer and more effective agrichemicals, pesticides, herbicides, and food additives have improved our food quantity, quality (*e.g.*, nutrition). Beneficial pharmaceutical agents have helped us achieve an unprecedented standard of health. Polymers and plastics affect our lives immensely. The more exotic materials of the nano-age are of growing importance. Chemical synthesis enabled the development of nearly every one of these beneficial commodities.

Webster's dictionary teaches that synthesis is "The process of producing a chemical compound (usually by the union of simpler chemical compounds)" (*1*). Complex molecules have been (bio)synthesized in nature for billions of years. But the biological pressures that drive natural evolution provide no incentive to improve beyond whatever is needed for short-term success, as determined by species survival. Humans, on the other hand, through their efforts during the last 100 years or so, have made tremendous advances, thereby turning synthesis into a science. This achievement enables today's synthetic chemists–through activities ranging from planned, reproducible laboratory exercises to impressive manufacturing successes–to drive and then deliver new discoveries as never before.

The field of "synthesis" is large. To make the task of summarizing its evolution somewhat more tractable, we have chosen to focus most of the discussion here on organic synthesis that is, the preparation of carbon-containing compounds.

The chapter is organized as follows. Some events from "Early History" in the field (*2*) are presented where, among other things, the role of serendipity is evident. The "Key Enabling Factors" that have propelled the evolution of the field are then discussed. Sections on "Synthesis of Pharmaceuticals and Other Bioactive Agents" and "Natural Product Synthesis" serve to demonstrate both the considerable impact and the tremendous growth of the field of organic synthesis. We close with some thoughts on "The What vs. How of Synthesis."

Early History

Organisms have been producing (biosynthesizing) complex organic compounds solely under the influence of natural forces since the beginning of life. However, starting with the accidental production of urea by Wöhler in 1828, human intervention has dramatically changed the pace and the spectrum of substances produced by that natural course. In the attempt to make ammonium cyanate (NH_4OCN) (*3,4*) by the combining of ammonium sulfate and potassium cyanate, urea (Figure 1) was formed instead. That this natural substance could be produced *in vitro* challenged the commonly held views of Vitalism, which attributed organic substances with a unique force. Berzelius, for example "asserted that a peculiar vital force intervenes in the formation of

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

organic compounds and their preparation in the laboratory can hardly be expected" (*5*).

In 1856 Perkin made an accidental discovery during unsuccessful attempts to synthesize the drug quinine. He obtained instead dark tars, one group of which was produced from the coal-tar component aniline. These substances (mauveines A-C, Figure 1) were found to be an excellent purple dye, which Perkin named mauve, and which he proceeded to manufacture (*6*). Commercial incentive inspired the efforts of others in the area of dye synthesis. The red pigment alizarin was produced from anthraquinone in 1869 (Figure 1, (*7,8*)). Indigo was prepared in 1878 by Baeyer (*9*) in six steps via isatin (Figure 1, (*10*)) and later, in 1890, from aniline via oxindole in three steps by Heumann (*11*).

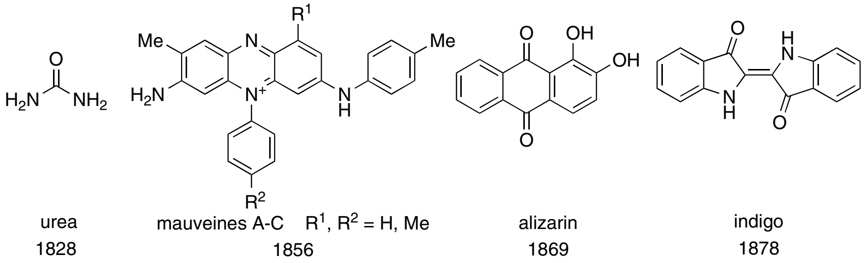


Figure 1. Synthetic chemistry accomplishments of the 19th century.

There are a number of examples of accidental discoveries that have had a major impact on polymer synthesis or derivatization. These include the Goodyear vulcanization of rubber (1840, (*12*)), an observation of Chardonnet that led to the discovery of rayon (1878, (*13*)), and the Dupont discoveries of the conversion of chloroprene to neoprene (1920s, (*14*)), the "cold drawing" method for making nylon (1937, (*15* , *16*)), and of polymerization of tetrafluoroethylene to Teflon® (1938, (*17*)).

Arguably, the first planned synthesis and, seemingly, the first use of the term synthesis in the sense that it is now employed is associated with Kolbe's multistep preparation of acetic acid (the active component of vinegar) from carbon in 1845 (Figure 2, (*18*)). In addition to the historical noteworthiness of this work, it is interesting to speculate whether it also has served as the inspiration of a common exercise still espoused in the introductory organic chemistry classroom. Many a novice chemist will recall being asked to provide a synthesis of target X starting from earth (C, Fe, S), air (O_2, H_2, Cl_2), fire (Δ), and water (H_2O)–or at least the moral equivalent.

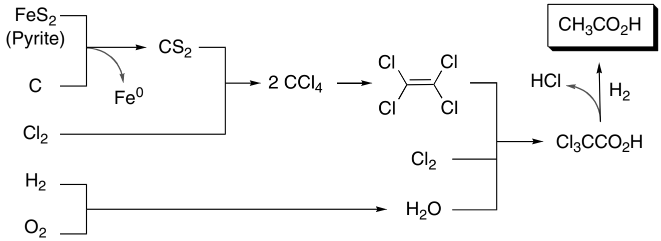


Figure 2. Kolbe's synthesis of acetic acid.

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

Key Enabling Factors

As has been noted elsewhere (*19*), various factors have played a key enabling role in the evolution of organic synthesis: 1) the development of spectroscopic (and other physical) methods for structural analysis; 2) the use of chromatographic methods of analysis and separation; 3) the delineation of structural limitations of organic compounds, particularly from the viewpoint of energetics; 4) the elucidation of the mechanisms of organic reactions; 5) the development of conformational analysis and its application to ground state and transition state structures; and 6) the need to discover new synthetic methodologies, often based on the development of new reagents and catalysts. This evolution has been bidirectional; most of these factors were themselves synergistically advanced by the synthesis of relevant probe compounds.

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

1) Spectroscopy

Prior to the era of spectroscopy, combustion analysis (*20*) (to establish molecular composition) and melting (or boiling) point characteristics (both of the substances themselves and, often, of covalent derivatives) were the mainstays of characterization of new compounds. The advent of various spectroscopic tools then greatly enabled the structure determination of new chemical substances. Widespread adoption of ultraviolet (UV, 1940s), infrared (IR, 1950s), and most significant of all, nuclear magnetic resonance (NMR, 1950s (*21*)) spectroscopies had a huge impact on the field. The advents of Fourier-transform (FT) methods and various two-dimensional correlation spectroscopies, routinely used today, came into play in the 1970s (*22*).

Similarly, mass spectrometry (MS) has had a dramatic impact, and is, arguably, the tool that continues to see the greatest pace of technological development. From its earliest days (*23*) the 'mass-spec' experiment has evolved from a substantial technological challenge to one that now can be almost trivial. One particularly important development that spurred this growth was the implementation of new ionization techniques (*24*), most notably that of electrospray ionization, which permits sample introduction at atmospheric pressure (*25*). The large number of instruments today that provide "walk-up" access to mass data, literally, at moments notice, even for the novice user attests to these advances as well as to the immense value of this tool.

2) Chromatography

Chromatographic separation technologies, in both the gas and liquid phases, have had a large impact on the science of synthesis. These are used in both analytical and preparative modes. The ability to isolate pure substances from multicomponent composites (*e.g.*, from crude reaction product mixtures) by methods other than classical recrystallization or distillation techniques greatly enhanced the types of research studies that can now be undertaken. It is safe to say that virtually none of today's research projects involving organic synthesis could be undertaken in the absence of these chromatographic methods.

Instruments for carrying out gas chromatography (GC), originally called gas-liquid chromatography (GLC) by its earlier developer Archer J. P. Martin (*26*) were first commercialized in the 1950s. Packed column technologies, which supported both analytical and preparative experiments, gave way to the use of capillary columns, which dominate the field today. While the latter generally give far more efficient separations because of their higher plate counts, an irony is that their low load capacity means that preparative GC protocols have largely disappeared from the modern lexicon of organic chemistry.

Martin was also largely responsible for the introduction of silica gel as a stationary phase for liquid partition chromatography (and of paper chromatography, which had a large impact on separations of biologicals). The popular introduction of reversed-phase supports in the 1970s launched their ascension to the place of dominance that they now enjoy, at least with respect to applications in HPLC analysis. To this day, however, normal phase separations on silica gel is the predominant method for preparative separations in synthetic chemistry laboratories. The complementary analytical tool of thin-layer chromatography (TLC), most commonly done on silica gel supports, was introduced by Stahl in the mid-1950s (*27*). Some sense of the importance of these techniques is revealed by the fact that the publication by W. Clark Still in 1978 first describing 'flash chromatography' (*28*) has been cited over 7500 times and is taught, in one form or another, to every practitioner of organic synthesis.

While not a mainstay of 'small molecule' synthetic chemistry studies, the techniques of ion exchange and, especially, size exclusion (*e.g.,* gel permeation) chromatographies have greatly impacted the analysis and purification of polymeric and biological samples.

The robust nature of some of these chromatographic techniques has allowed them to be used in tandem with some spectroscopic methods, especially mass spectrometry. This has given rise to the now routine hyphenated techniques of GC-MS and LC-MS. The fact that it is not uncommon for modern organic synthesis laboratories to have an LC-MS instrument per every 8-12 bench chemists attests to the power and nearly indispensable value of the technique.

As a final aside, we note that technological advances, even when they address a somewhat mundane operation, can have considerable practical impact. In this vein, it is interesting to note the 1950 paper "Versatile Laboratory Concentration Device" (a.k.a. rotary evaporator), consisting of one Figure and two paragraphs (*29*), and the 1944 patent "Magnetic Stirrer" (*30*).

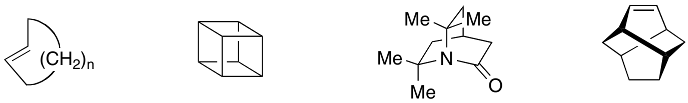
3) Structure

The power of synthetic chemistry has enabled chemists to address fundamental questions of structure (and mechanism, see below). A notable early example is the first synthesis and resolution of a chiral compound containing only one carbon atom (chloroiodomethane sulfonic acid), which was reported in 1914 (Pope, (*31*)). Prior to that many were of the opinion that molecules

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

comprising optically active samples needed to contain a skeleton with at least two carbon atoms. Subsequent studies, some represented by the structures in Figure 3, addressed questions of structural feasibility through the synthesis of so-called "theoretical" (a.k.a. non-natural) molecules. Examples (Figure 3) of these non-conventional targets include compounds that have allowed chemists to probe questions of strain, like *trans*-cycloalkenes (Cope, (*32*)), cubane (Eaton, (*33*)), bridgehead amides (e.g., azabicyclo[2.2.2]octan-2-one, Pracejus, (*34*) and 2-quinuclidone, Stoltz, (*35*)), twisted/nonplanar alkenes (e.g., twistene, Tichy, (*36*)), "anti-Bredt" bridgehead olefins (bicyclo[3.3.1]non-1-ene, Marshall, (*37*); Wiseman (*38*); bicyclo[n.1.0]alk-1-enes, Köbrich, (*39*)), cycloalkynes (Blomquist, (*40*)), [1.1.1]propellane (Wiberg, (*41*)), and trans-fused bicyclic compounds (Paukstelis, (*42*); Gassman, (*43*)).

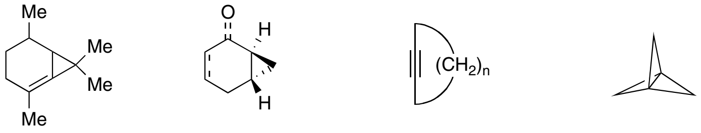


trans-cycloalkenes
E-cyclooctene (n=4)
1953

cubane
1964

bridgehead amides
1-azabicyclo[2.2.2]-octan-2-one
1965

twistene
1969



anti-Bredt alkenes
[4.1.0]hept-1-ene
1972

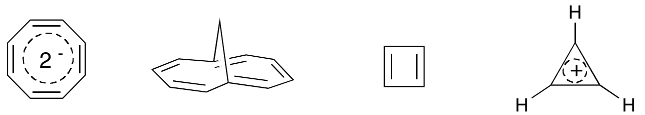
trans-fused
cyclopropanes
1972, 1983

cycloalkynes
cyclooctyne (n=4)
1973

[1.1.1]propellane
1982

Figure 3. Molecules synthesized to address questions of strain.

Aspects of electronic theory like aromaticity (Figure 4) (and studies of orbital symmetry control, see later) have been probed through the synthesis of molecules like cyclobutadiene (*44* , *45*), bridged annulenes (*46*), the cyclopropenium cation (*47*), and the cyclooctadiene dianion (*48*).



cyclooctadiene
dianion
1960

1,6-methano[10]-
annulene
1965

cyclobutadiene
1965, 1973

cyclopropenyl
cation
1967

Figure 4. Molecules synthesized to address questions of aromaticity.

Synthesis has enabled chemists to explore new topological space, as represented in Figure 5 by the catenanes (*49*), rotaxanes (*50*), and molecules with a Möbius topology (*51*).

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

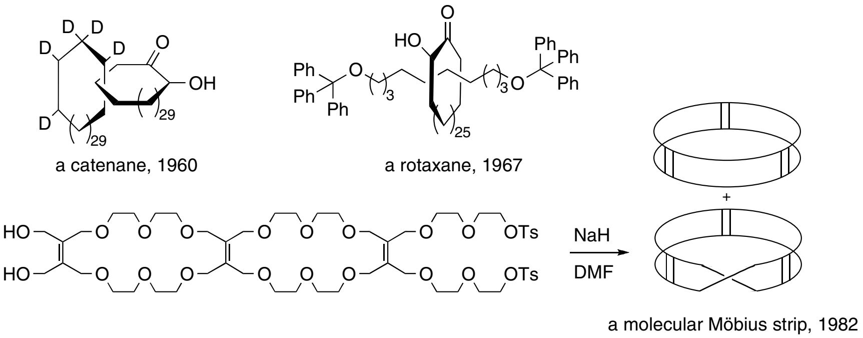


Figure 5. Molecules synthesized to address questions of unusual topology.

Many discoveries in synthesis have their own anecdotal histories, one of which—the first synthesis of an organometallic metallocene, ferrocene—is mentioned here. In 1951, the newly appointed Assistant Professor Peter L. Pauson, working with an undergraduate student, Thomas J. Kealy, attempted to oxidatively couple cyclopentadienyl magnesium bromide with ferric chloride (*52,53*). The goal was to prepare fulvalene to test the then recent proposal (*54*) that it would display aromatic character. Instead, a very stable orange solid was produced, which proved to be the famous ferrocene, about whose structure determination much has been written (Figure 6, (*55,56,57*)).

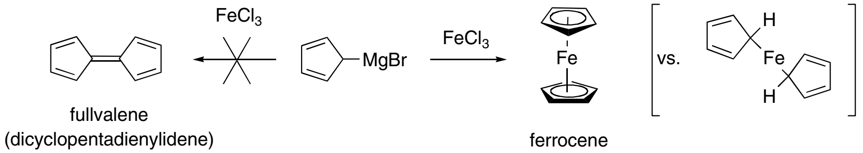


Figure 6. Unintended synthesis of ferrocene.

4) Mechanism

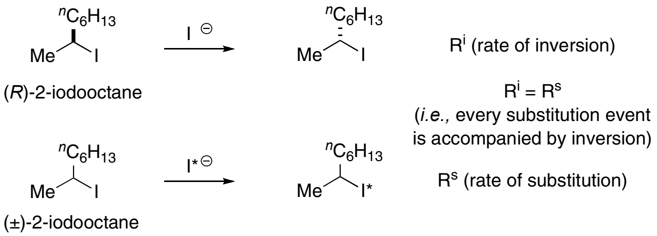
Mechanistic understanding is a fundamental underpinning of organic chemistry. Mechanism and synthesis have always benefited from their symbiotic relationship. Mechanistic reasoning empowers researchers as they conceive and plan new objectives in synthesis.

Sir Robert Robinson played an early central role in the development of the electronic theory of organic chemistry (*58*). For example, he was the first to use the now commonplace curly arrow to imply the reorganization of electron density during the course of a chemical reaction (*59*). The construct of stereoelectronic control (which, in its most literal sense, explains *all* chemical reactions) underpins many concepts now in the lexicon of mechanistic and synthetic organic chemistry. These include: the Hammond postulate, the Curtin-Hammett Principle, the Markovnikov rule (for additions to alkenes), the Thorpe-Ingold effect (on rates of cyclization), the Bürgi-Dunitz approach trajectory, Cram/Cornforth/Felkin-Ahn controlled additions (to chiral ketones

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

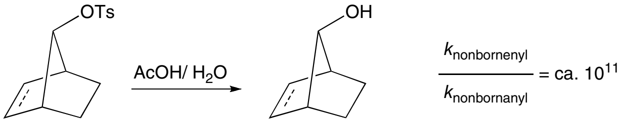
and aldehydes), the Zimmerman-Traxler transition state model (for aldol addition reactions), and Baldwin's rules (for ring closure).

Likewise, reliance on the power of synthesis has allowed chemists to conceive experiments that address crucial mechanistic questions and to then prepare compounds for the execution of those studies. One early and most notable example is the optically active 2-iodooctane that Hughes *et al.* used in elegant work to establish that the S_N2 reaction proceeds with inversion of configuration (Figure 7, (*60*)). Other examples are the bicyclo[2.2.1]heptanyl and –enyl tosylates used in solvolysis studies by Saul Winstein that advanced the principle of neighboring group participation (Figure 8, (*61*)) and the *E*- and *Z*-hexadeutero-2,3-dimethyl-2-butenes with which L. M. Stephenson established the intermediacy of the zwitterionic perepoxide X in the net ene reaction of alkenes with singlet oxygen (Figure 9, (*62*)).



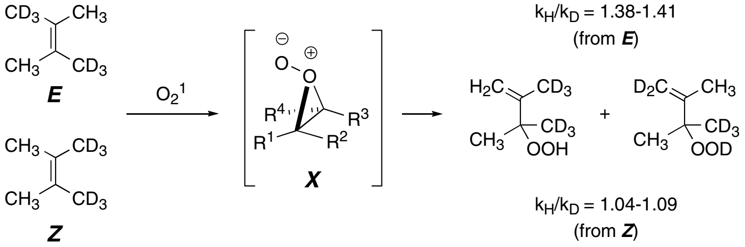
$^nC_6H_{13}$
Me
I
$I^\ominus$
(*R*)-2-iodooctane
R^i (rate of inversion)
$R^i = R^s$
(*i.e.*, every substitution event is accompanied by inversion)
R^s (rate of substitution)
$I^{*\ominus}$
I*
(±)-2-iodooctane

Figure 7. Studies of the stereochemical course of the S_N2 reaction.



OTs
AcOH/ H_2O
OH
$\frac{k_{\text{nonbornenyl}}}{k_{\text{nonbornanyl}}}$ = ca. 10^{11}

Figure 8. Studies of neighboring group participation and proximity acceleration.



CD_3
CH_3
E
Z
O_2^1
X
R^1
R^2
R^3
R^4
k_H/k_D = 1.38-1.41 (from *E*)
H_2C
OOH
+
D_2C
OOD
k_H/k_D = 1.04-1.09 (from *Z*)

Figure 9. Studies of the mechanism of the singlet oxygen ene-reaction.

Bullvalene (Figure 10) is a molecule that was predicted to undergo rapid degenerate Cope rearrangement (*63*). This hypothesis was only testable once the synthesis of bullvalene had been achieved (*64*). "At 100 °C the NMR-spectrum of bullvalene exhibits a most remarkable property, *viz.* only one sharp resonance $\tau = 5.8$. On cooling, this band broadens. At -25 °C, two bands are apparent, one centered at $\tau = 4.0$ with relative area 6, the other centered at $\tau = 7.9$ with a relative area of 4." Semibullvalene has analogous properties (*65*).

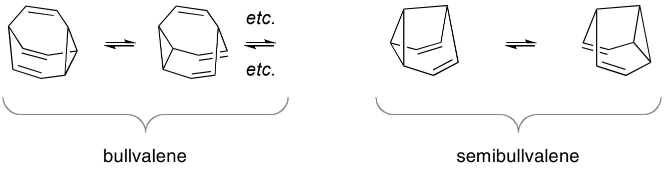


Figure 10. Studies of facile degenerate rearrangements.

Finally, certain theories and rules have evolved both to explain and predict various reaction outcomes. These have often benefited, again in synergistic and iterative fashion, from the synthesis of specific compounds designed to further test and delineate the growing theory. Frontier molecular orbital (perturbation) theory (*66,67,68*) and the theory of conservation of orbital symmetry (*69,70*)–the Woodward-Hoffman rules–are two examples that contribute substantially to the foundation of modern mechanistic thinking (*71,72*). The octatrienes shown in Figure 11 are just one (of many) example(s) of such probe molecules, used in this case to validate the rules for electrocyclic cyclization reactions (*73*).

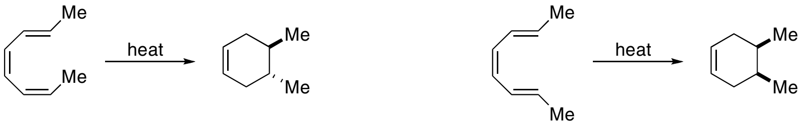


Figure 11. Octatrienes used to support electrocyclic cyclization rules.

5) Conformational analysis

Ideas about molecular geometry can be traced to the earliest of fundamental theories about structure and bonding in organic compounds (*e.g.,* van't Hoff (*74,75*) and Baeyer (*9,76*)). However, the great importance of the analysis of conformation—the different geometries that a molecule can adopt simply by rotation about bonds—is a more recent development. Many would say this growth started with the analysis of cyclohexyl ring systems by Sir Derek H. R. Barton and Odd Hassel (*77*) (e.g., Figure 12, panel **a** (*78*)). Newer yet, concepts such as anomeric stabilization (which govern reactions like the equilibrium shown in panel **b** of Figure 12), stereoelectronic effects (which explain geometry-based differences in reactivity, such as the faster hydrolysis of the cis-fused acetal in panel **b** (*79*)), and allylic strain (which, e.g., allows prediction of the preferred side (face) of approach by reagents to a simple chiral alkene like the one shown in panel **c** (*80,81*)) are invaluable for the understanding and planning of modern synthetic chemistry research.

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

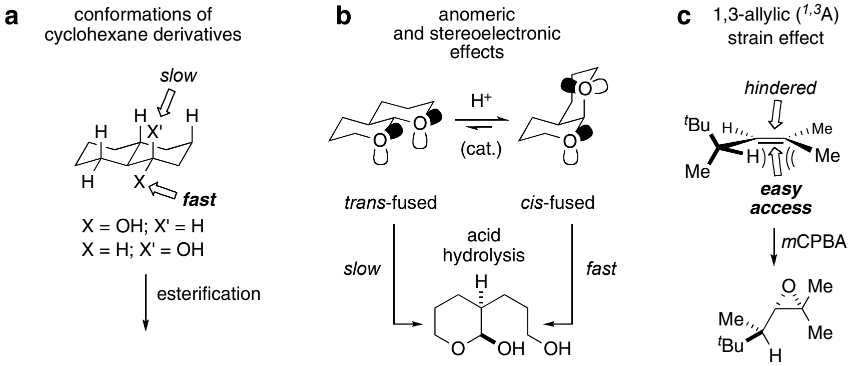

a
conformations of cyclohexane derivatives
slow
fast
X = OH; X' = H
X = H; X' = OH
esterification
b
anomeric and stereoelectronic effects
H+
(cat.)
trans-fused
cis-fused
acid hydrolysis
slow
fast
c
1,3-allylic (1,3A) strain effect
hindered
easy access
mCPBA

Figure 12. Examples of important concepts in conformational analysis.

6) Synthetic Methodologies: Often via New Reagents and Catalysts

Synthetic methods are processes having sufficient generality and efficiency that they can be reliably used in synthesis. The literature is replete with examples, from esoteric to highly general in applicability and from robust to finicky in their implementation. We let Nobel citations guide our choice of those to mention here.

The examples shown in Figure 13 involve the use of stoichiometric reagents:

- 1912, Victor Grignard "for the discovery of the so-called Grignard reagent"
- 1947, Sir Robert Robinson, developer of the Robinson annulation reaction, "for his investigations on plant products of biological importance, especially the alkaloids."
- 1950, Otto Diels and Kurt Alder “for their discovery and development of the diene synthesis."
- 1979, Herbert C. Brown and Georg Wittig "for their development of the use of boron- and phosphorus-containing compounds, respectively, into important reagents in organic synthesis."
- 1984 Bruce Merrifield "for his development of methodology for chemical synthesis on a solid matrix" (*82*).
- 1990, Elias J. Corey "for his development of the theory and methodology of organic synthesis" (silyl ether protecting group strategies are one example (*83*).

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

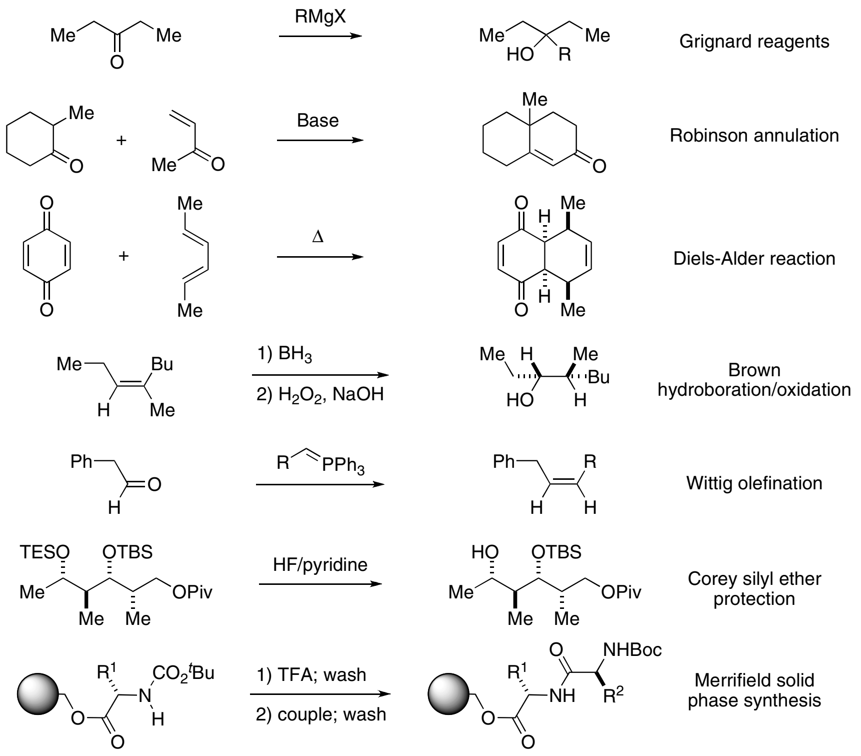

RMgX
Grignard reagents
Base
Robinson annulation
Δ
Diels-Alder reaction
1) BH_3
2) H_2O_2, NaOH
Brown hydroboration/oxidation
Wittig olefination
HF/pyridine
Corey silyl ether protection
1) TFA; wash
2) couple; wash
Merrifield solid phase synthesis

Figure 13. Examples of Nobel synthetic methodologies.

The examples shown in Figure 14 are methods based on the use of metal-containing catalysts:

- 1912, Paul Sabatier "for his method of hydrogenating organic compounds in the presence of finely disintegrated metals."
- 2001, William S. Knowles, Ryoji Noyori "for their work on chirally catalyzed hydrogenation reactions" and K. Barry Sharpless "for his work on chirally catalyzed oxidation reactions."
- 2005, Yves Chauvin, Robert H. Grubbs, and Richard R. Schrock "for the development of the metathesis method in organic synthesis."

Methods for asymmetric catalysis are additionally valuable because of the requirement that chiral pharmaceutical agents be sold in enantiomerically pure form.

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

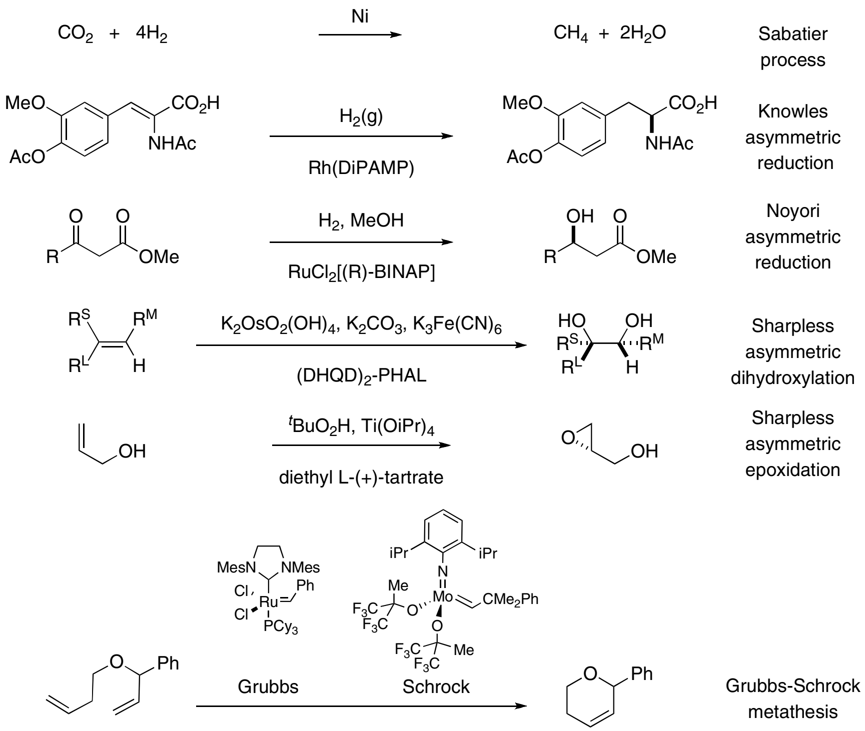

CO2 + 4H2
Ni
CH4 + 2H2O
Sabatier process
MeO
AcO
CO2H
NHAc
H2(g)
Rh(DiPAMP)
Knowles asymmetric reduction
O
R
OMe
H2, MeOH
RuCl2[(R)-BINAP]
OH
Noyori asymmetric reduction
RS
RM
RL
H
K2OsO2(OH)4, K2CO3, K3Fe(CN)6
(DHQD)2-PHAL
HO
Sharpless asymmetric dihydroxylation
tBuO2H, Ti(OiPr)4
diethyl L-(+)-tartrate
Sharpless asymmetric epoxidation
MesN
NMes
Cl
Ru
Ph
PCy3
iPr
N
Mo
Me
F3C
CMe2Ph
Grubbs
Schrock
Grubbs-Schrock metathesis

Figure 14. Examples of Nobel metal-catalyzed synthetic methodologies.

As with the ferrocene synthesis mentioned earlier, anecdotal accounts often reveal a human component that adds to the intrigue of discovery. The so-called Huang Minlon modification of the Wolff-Kishner method for reduction of ketones to alkanes (as recounted to one of the authors by Mary Fieser) is no exception. In the late 1940s Huang Minlon, working in the laboratory of Professor Louis F. Fieser at Harvard University, was exploring a Wolff-Kishner reduction that required an extended reflux period of several days. Huang decided to take an out of town trip for the weekend. Upon his return, he noted that the condenser and reaction flask had separated but, importantly, that the experiment had produced a substantially higher yield than in earlier attempts. He speculated that loss of water and excess hydrazine had, in effect, allowed the reaction mixture to reach a substantially higher temperature, thereby facilitating the rate-limiting breakdown of the hydrazone intermediate (Figure 15). Subsequent experiments quickly supported that hypothesis, leading to a protocol in which ethyleneglycol (bp, 197 °C) was used as the reaction solvent, the temperature of the reaction vessel was raised to ca. 195 °C, and the water and excess hydrazine were intentionally removed by distillation (*84*). The rest is history.

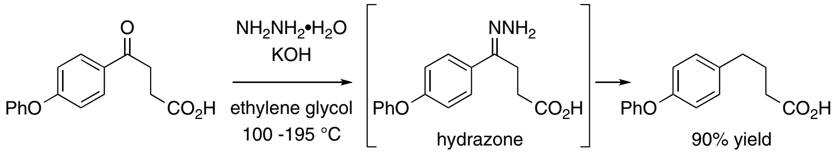

O
PhO
CO2H
NH2NH2•H2O
KOH
ethylene glycol
100 -195 °C
NNH2
hydrazone
90% yield

Figure 15. Huang-Minlon modification of the Wolff-Kishner reduction.

Synthesis of Pharmaceuticals and Other Bioactive Agents

Synthesis has enabled, for better or worse, the production of compounds capable of producing powerful biological responses in living organisms. Early on, these were usually natural compounds or derivatives thereof. Notable examples are the myriad of folkloric agents, the β-lactam antibiotics (*e.g.,* cephams and penams), steroid hormones, and paclitaxel (see below).

Perhaps it was inevitable, but synthesis has also empowered those with unprincipled motivations. Illicit drugs, agents for biological weaponry, and designer steroids, although different in the magnitude and focus of their impact, all serve to human detriment (Figure 16).

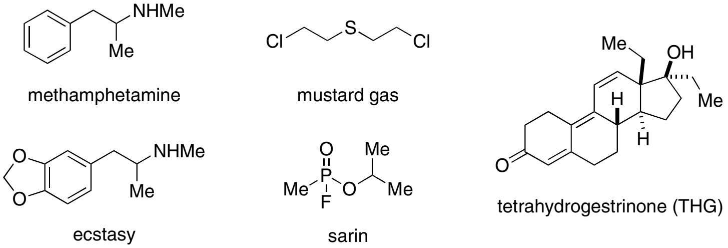


Figure 16. Examples of sinister agents accessible by chemical synthesis.

From a brighter perspective, synthetic chemistry also has provided access to a large number of beneficial medicinal agents. The majority of these regulated pharmaceuticals that have been brought to market during the era of modern medicines are non-natural products. All have been (or are) synthesized. The earliest examples of such drugs are the anti-syphilitic/anti-trypanosomiasitic agent arsphenamine (trade named Salvarsan) (*85*) discovered in 1910 by Ehrlich and the broad spectrum antibiotic named Prontosil, the prodrug precursor to sulfanilamide, developed at Bayer Laboratories in the 1930s (Figure 17).

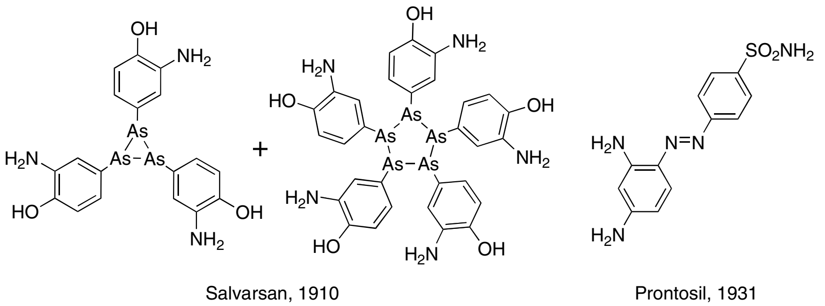


Figure 17. Early examples of synthetic pharmaceutical agents.

The drug paclitaxel (Taxol®, Figure 18) posed a different kind of challenge for synthetic chemistry. This natural product possesses many beneficial anti-cancer properties, but the compound can only be isolated from the bark of the Pacific Yew tree. This slow growing plant must be sacrificed to harvest its paclitaxel. It was clear that demand could not be met if this were the only source. While an economically viable total synthesis of this complex structure

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

has not yet been achieved, it was discovered that a related compound, 10-deacetylbaccatin III, can be isolated from a renewable source, the needles of a related European Yew (*Taxus baccata*). Synthetic chemists in the laboratory of Robert A. Holton established that paclitaxel could be synthesized on a production scale via a semisynthetic route (*86*).

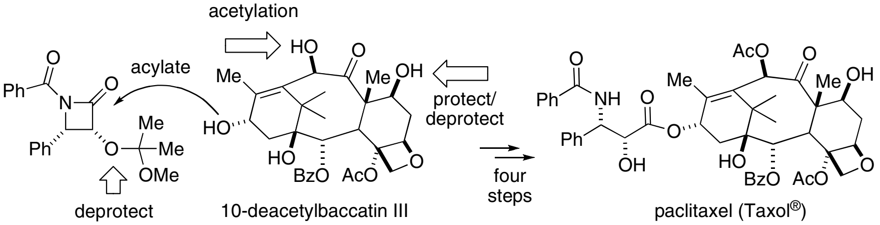

acetylation
acylate
protect/ deprotect
four steps
deprotect
10-deacetylbaccatin III
paclitaxel (Taxol®)

Figure 18. Route for the large-scale production of the anticancer drug Taxol®.

Natural Product Synthesis

Some Milestone Accomplishments

Natural product structures have inspired synthetic chemists from the onset of activities that comprise the discipline of organic synthesis. The origin of many innovative developments in the field can be traced to problems and challenges raised by the molecules of nature and to the thoughts and experiments that ensued in the course of meeting those challenges.

Urea and acetic acid notwithstanding (see above), the era of natural product synthesis began roughly near the start of the 20th century. Among the most noteworthy early accomplishments are the syntheses of the targets shown in Figure 19 (glucose, E. Fischer (*87,88*); camphor, Komppa (*89*) and Perkin (*90*); α-terpineol, Perkin (*91*); tropinone, Robinson (*92,93*); haemin, H. Fischer (*94,95*); equilenin, Bachmann (*96*); and quinine, Rabe (*97*) and Woodward (*98,99,100,101*)).

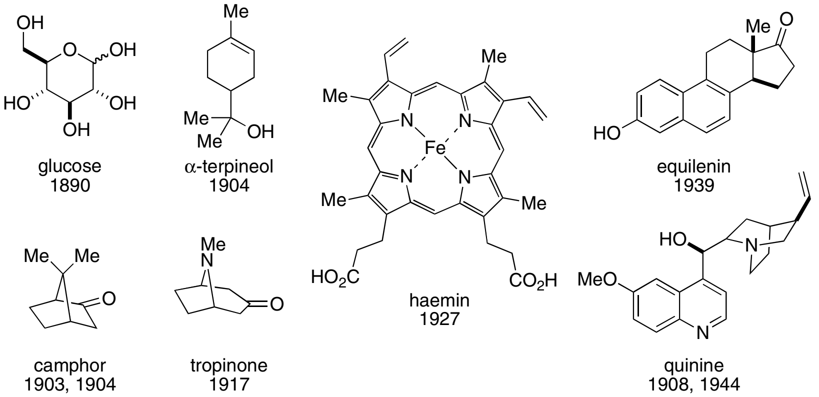

glucose
1890
α-terpineol
1904
haemin
1927
equilenin
1939
camphor
1903, 1904
tropinone
1917
quinine
1908, 1944

Figure 19. Natural product targets from the early 20th century.

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

As reflected by the structures in Figure 20, the pace of synthesis of molecules of greater structural complexity and/or biological importance increased in the post-war efforts of the 1950s. These targets include cortisone (Woodward (*102*)), morphine (Gates (*103*)), cedrene (Stork (*104*)), reserpine (Woodward (*105*)), and penicillin V (Sheehan (*106*)).

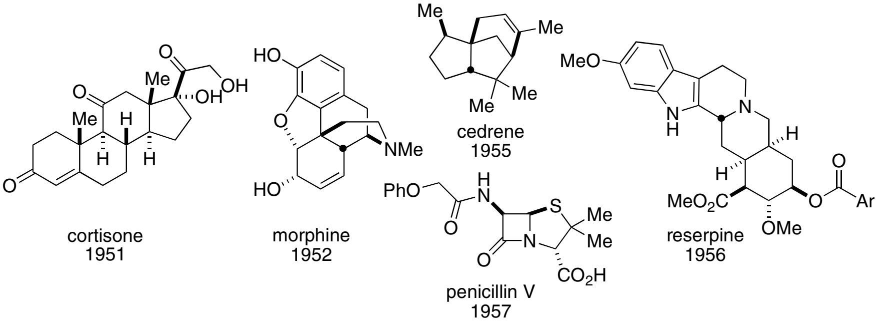


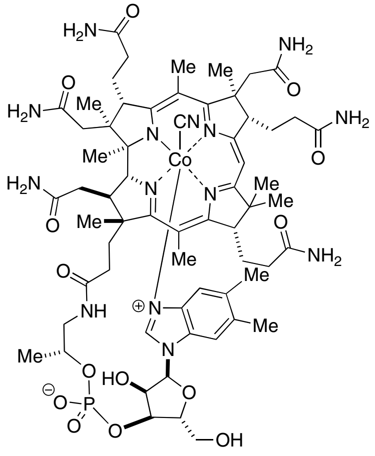
Figure 20. Natural product targets from the 1950s.

Total syntheses of yet more complex structures (benefiting, no doubt, to an ever increasing extent from the logic and formalism of retrosynthetic analysis (*19*)) served to demonstrate both the growing level of sophistication of the discipline as well as the remarkable abilities–both intellectual and experimental–of its practitioners. These include (Figure 21) molecules like vitamin B_{12} (Woodward and Eschenmoser (*107,108,109,110*)), erythronolide B (Corey, 1978, (*111,112*)), dodecahedrane (Paquette, 1982, (*113,114,115*)), and palytoxin (Kishi, 1994, (*116,117,118,119*)). The last of these emphasizes, in a most dramatic fashion, a time-honored role (cf. glucose (*88*)) that synthesis has played in the field of natural products chemistry–namely as an irrefutable proof of structure. Especially important is the incisive role that synthesis continues to play in the correction of erroneously reported natural product structures, as demonstrated by the recent examples of hexacyclinol ("Occasionally, blatantly wrong science is published (*120*), and to the credit of synthetic chemistry, the corrections usually come quickly and cleanly (*121,122*)" E. J. Corey in (*123*)), neopeltolide (*124,125*) and palmerolide (*126,127*).

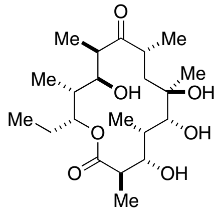
Biomimetic Synthesis

Natural product structures often elicit biomimetic reasoning that leads to mechanistic hypotheses for new reaction processes. Two notable examples of syntheses that capitalized on such thinking are the progesterone (*128*) and homosecodaphniphyllate (*129*) syntheses highlighted in Figure 22. The cation-olefin cyclization used by William S. Johnson drew its inspiration from the Stork-Eschenmoser hypothesis (*130,131,132*) for enzymatic squalene

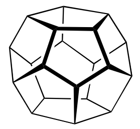
Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010



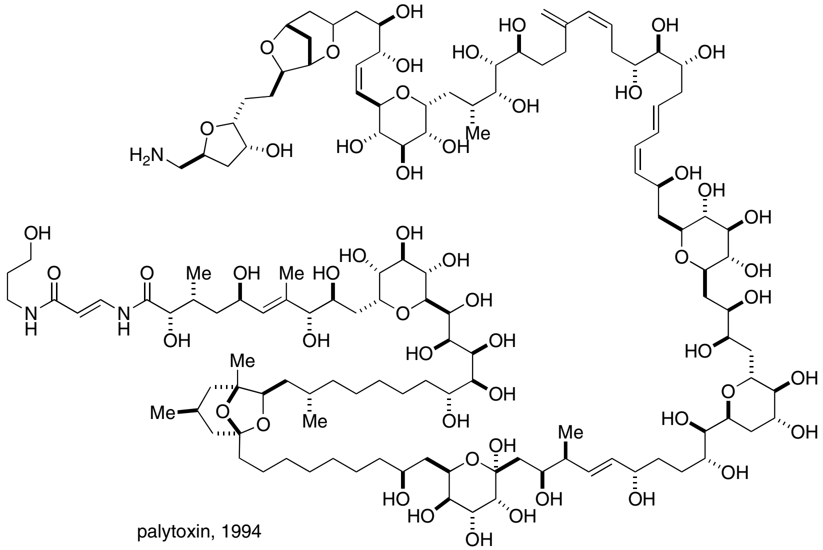
vitamin B$_{12}$, 1973



erythronolide B, 1978



dodecahedrane, 1982

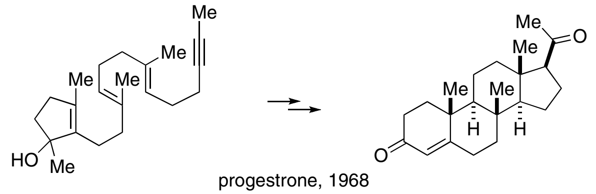


palytoxin, 1994

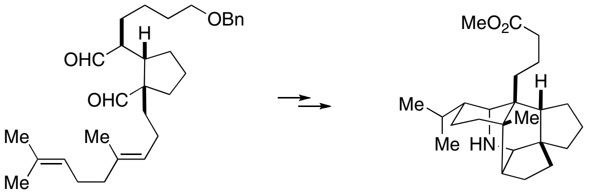
Figure 21. Examples of synthesized molecules having high structural complexity.

cyclization in the biosynthesis of steroids. Clayton H. Heathcock established the entire structural framework of the alkaloid via a one-pot cascade of sequential cyclizations.

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010



progesterone, 1968



homosecodaphniphyllate, 1992

Figure 22. Examples of biomimetic syntheses.

The field of natural product synthesis remains vibrant. Objective evidence can be seen in Table I, in which the number of journal articles containing the words "total synthesis" in the publication title are listed by decade.

Table I. Number of Publications with "Total Synthesis" in the Title.

Decade	*All Journals*[a]	*ACS Journals*[b]	*Decade*	*All Journals*[a]	*ACS Journals*[b]
1901-1910	5	0	1951-1960	325	91
1911-1920	6	0	1961-1970	1198	224
1921-1930	5	0	1971-1980	2835	468
1931-1940	40	5	1981-1990	6148	880
1941-1950	85	15	1991-2000	6380	1190
			2001-March 2008	5520	1364

[a] From a SciFinder search. [b] From an Advanced Article Search at ACS Publications.

As leading practitioners Professors K. C. Nicolaou and Erik J. Sorensen have stated, the science of natural product synthesis continues to be motivated, in part, "by the need to improve our ability to synthesize organic molecules in more efficient and economical ways (*e.g.*, through green (*133*) or atom economical (*134*) chemistries (often catalytic)) … these driving forces are likely to persist for some time to come" (*135*).

The 'What *vs.* How' of Synthesis

We close with some thoughts about the current state of affairs in the field of synthesis. The primary driver today of *why* synthesis is performed falls largely into one of two camps. Either the significance lies in *'what'* is being made or in *'how'* the target is being made. There is very little overlap, which, it seems to us, is both understandable and as it should be. Neither type of endeavor is more important or loftier than the other–the two are just different.

Projects that are motivated by ‘what’, are very often done in a collaborative, multidisciplinary, or team setting. The overall objective typically is to produce a new chemical entity with a certain property or function. This often requires synthesis both at the outset as well as iteratively throughout the course of the overall study. The choice of target for each subsequent round of synthesis is often informed by feedback arising from a measured function or property of the product of the previous round. This approach is very common in medicinal chemistry laboratories in the course of drug discovery projects as well as in the production of new materials comprised of, *e.g.*, synthetic polymers. Obviously, synthesis can drive many additional types of studies in the fields of mechanistic chemistry (see above), catalysis (*e.g.,* ligand synthesis or metal complex synthesis), medicine (*e.g.,* imaging agents), or biology (*e.g.,* molecular probes or cell cycle regulation agents). In *every* case, if the synthetic chemist cannot put the ‘what’ into a bottle or vial, the collaborator has nothing to study. An important consequence is that the choice of synthesis plan in these settings is nearly always governed by practicality; a corollary is that highly innovative (and, necessarily, more risky) approaches are rarely the most prudent.

In contrast, projects that are driven by ‘how’ the target is made need not be, and, in our opinion, should not be bound by ‘what’ is being made. This provides more ‘target structure space’ within which to innovate. Highly creative thinking is, by definition, a rare commodity, so why artificially impose constraints of subsets of targets? Like it or not, the outcomes of creative ideas and of the experiments they precipitate are then judged on a ‘wish I had thought of that’ basis, in much the same way as the output from performers (‘wish I could do that’) and artists (‘wish I had done that’) are appreciated. As said above, neither of the what- vs. how-driven synthesis activity is inherently superior or more significant–the two are just different.

Conclusion

Professor Barry M. Trost has provided useful definitions of various kinds of *selectivity* that apply to nearly all chemical transformations (*136*). The ideal reaction is perfectly *chemoselective* (“performs a wanted structural change and none other”), *regioselective* (“orients the reacting partners in correct fashion”), *diastereoselective* (“creates the correct orientations of the various parts of the molecule with respect to each other”), and *enantioselective* (“enables the formation of a molecule of one-handedness or a mirror image isomer”). None yet meets all of these goals. Considerable challenges and opportunities remain for practitioners of synthesis.

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

References

1. http://www.websters-online-dictionary.org, 1st definition of the noun "synthesis" (accessed July 19, 2008).
2. Hudlicky, T.; Reed, J. W. *The Way of Synthesis. Evolution of Design and Methods for Natural Products;* WILEY-VCH Verlag GmbH & Co: Weinheim, 2007.
3. Wöhler, F. *Ann. Phys. Chem.* **1828,** *12,* 253–256.
4. Cohen, P. S.; Cohen, S. M. *J. Chem. Educ.*, **1996,** *73,* 883–886.
5. Parrington, J. R. *A History of Chemistry;* MacMillan and Co Ltd.: London, 1964; Vol. 4, pp 142–177.
6. Brock, W. H. *The Norton History of Chemistry;* W. W. Norton: New York, 1993; pp 297–301.
7. Gräbe, C.; Liebermann, C. *Ber. Dtsch. Chem. Ges.* **1869,** *2,* 332–334.
8. Caro, H.; Gräbe, C.; Liebermann, C. *Ber. Dtsch. Chem. Ges.* **1870,** *3,* 359–360.
9. Nobel Prize in Chemistry 1905: to Johann Friedrich Wilhelm Adolf von Baeyer "in recognition of his … work on organic dyes and hydroaromatic compounds." This partial quotation as well as the analogous ones in later references and text are taken from the citation synopses of the Nobel Foundation and are used with the permission of the Nobel Foundation.
10. Baeyer, A. *Ber. Dtsch. Chem. Ges.* **1878,** *11,* 1296–1297.
11. Heumann, K. *Ber. Dtsch. Chem. Ges.* **1890,** *23,* 3431–3435.
12. Pierce, B. *Trials of an Inventor: Life and Discoveries of Charles Goodyear;* Carlton and Porter: New York, 1866.
13. Roberts, R. M. *Serendipity: Accidental Discoveries in Science;* John Wiley & Sons, Inc.: New York, 1989; pp 94–96.
14. Berenbaum, M. B. *Encyclopedia of Polymer Science and Technology;* John Wiley & Sons, Inc.: New York, 1969; Vol 11, pp 425–447.
15. Roberts, R. M. *Op. cit.*; pp 170–176.
16. Carothers, W. H. U.S. Patent 2,071,250, 1937.
17. Roberts, R. M. *Op. cit.*; pp 187–191.
18. Kolbe, H. *Annalen* **1845,** *54,* 145–146.
19. Corey, E. J.; Cheng, X.-M. *The Logic of Chemical Synthesis;* Wiley: New York,1989; p. 3.
20. Nobel Prize in Chemistry 1923: to Fritzs Pregl "for his invention of the method of micro-analysis of organic substances."
21. Nobel Prize in Physics 1952: to Felix Bloch and Edward M. Purcell "for their development of new methods for nuclear magnetic precision measurements."
22. Nobel Prize in Chemistry 1991: to Richard R. Ernst for "the development of the methodology of high resolution NMR spectroscopy."
23. Beynon, J. H. *Mikrochim. Acta* **1956,** *44,* 437–453.
24. Nobel Prize in Chemistry 2002: to John B. Fenn and Koichi Tanaka (and Kurt Wüthrich for NMR "structure analyses of biological macromolecules") "for their development of soft desorption ionization methods for mass spectrometry."

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

25. Fenn, J. B. *J. Biomol. Techniques* **2002,** *13,* 101–118.
26. Nobel Prize in Chemistry1952: to Archer J. P. Martin and Richard L. M. Synge “for their discovery of partition chromatography.”
27. Stahl, E.; Schroter, G.; Kraft, G.; Renz, R. *Pharmazie* **1956,** *11,* 633–637.
28. Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978,** *43,* 2923–2925.
29. Craig, L. C.; Gregory, J. D.; Hausmann, W. *Anal. Chem*. **1950,** *22,* 1462.
30. Rosinger, A. U.S. Patent 2,350,534, 1944.
31. Pope, W. J.; Read, J. *J. Chem. Soc.* **1914,** *105,* 811–821.
32. Cope, A. C,; Pike, R. A.; Spencer, C. F. *J. Am. Chem. Soc*. **1953,** *75,* 3212–3215.
33. Eaton, P. E.; Cole, T.W., Jr. *J. Am. Chem. Soc.* **1964,** *86,* 962–964.
34. Pracejus, H.; Kehlen, M.; Kehlen, H.; Matschiner, H. *Tetrahedron* **1965,** *21,* 2257–2270.
35. Tani, K.; Stoltz, B. M. *Nature* **2006,** *441,* 731–734.
36. Tichy, M.; Sicher, J. *Tetrahedron Lett.* **1969,** *10,* 4609–4613.
37. Marshall, J. A.; Faubl, H. *J. Am. Chem. Soc.* **1970,** *92,* 948–955.
38. Wiseman, J. R.; Fletcher, W. A. *J. Am. Chem. Soc.* **1970,** *92,* 956–962.
39. Köbrich, G.; Baumann; M. *Angew. Chem. Int. Ed. Engl.* **1972,** *11,* 52–53.
40. Blomquist, A. T.; Liu, L. H. *J. Am. Chem. Soc*. **1953,** *75,* 2153–2154.
41. Wiberg, K. B.; Walker, F. H. *J. Am. Chem. Soc.* **1982,** *104,* 5239–5240.
42. Paukstelis, J. V.; Kao, J. *J. Am Chem. Soc.* **1972,** 94, 4783–4784.
43. Gassman, P. G.; Bonser, S. M. *J. Am. Chem. Soc.* **1983,** *105,* 667–669.
44. Emerson, E. W.; Pettit, R. *J. Am. Chem. Soc*. **1965,** *87,* 131–133.
45. Chapman, O. L.; McIntosh, C. L.; Pacansky, J. *J. Am. Chem. Soc.* **1973,** *95,* 614–617.
46. Gerson, F.; Heilbronner, E.; Böll, W. A.; Vogel, E. *Helv. Chim. Acta*, **1965,** *48,* 1494–1512.
47. Breslow, R.; Groves, J. T.; Ryan, G. *J. Am. Chem. Soc*. **1967,** *89,* 5048–5049.
48. Katz, T. J. *J. Am. Chem. Soc*. **1960,** *82,* 3784–3785.
49. Wasserman, E. *J. Am. Chem. Soc.* **1960,** *82,* 4433–4434.
50. Harrison, I. T.; Harrison, S. *J. Am. Chem. Soc.* **1967,** *89,* 5723–5724.
51. Walba, D.; Richards, R. M.; Haltiwanger, R. C. *J. Am. Chem. Soc*. **1982,** *104,* 3221–3223.
52. Kealy, T. J.; Pauson, P. L. *Nature* **1951,** *168,* 1039–1040.
53. Pauson, P. L. *J. Organometal. Chem.* **2001,** *637– 639,* 3–6.
54. Brown R. D. *Nature* **1950,** *165,* 566–567.
55. Zydowsky, T. M. *Chem. Intelligencer* **2000,** 29–34.
56. Laszlo, P.; Hoffmann, R. *Angew. Chem. Int. Ed. Engl.* **2000,** *39,* 123–124.
57. Wilkinson, G.; Rosenblum, M.; Whiting, M. C.; Woodward, R. B. *J. Am. Chem. Soc*. **1952,** *74,* 2125–2126.
58. Saltzman, M. D. *Nat. Prod. Reports* **1987,** *4,* 53–60.
59. Kermack, W. O.; Robinson, R. *J. Chem. Soc*. **1922,** *121,* 427–440.
60. Hughes, E. D.; Juliusburger, F.; Masterman, S.; Topley, B.; Weiss, J. *J. Chem. Soc.* **1935,** 1525–1529.
61. Winstein, S.; Shatavsky, M.; Norton, C.; Woodward, R. B. *J. Am. Chem. Soc*. **1955,** *77,* 4183–4184.

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

62. Grdina, S. B.; Orfanopoulos, M.; Stephenson, L. M. *J. Am. Chem. Soc.* **1979,** *101,* 3111–3112.
63. Doering, W. v. E.; Roth, W. R. *Angew. Chem. Int. Ed. Engl.* **1963,** *2,* 115–122.
64. Schröder, G. *Angew. Chem. Int. Ed. Engl.* **1963,** *2,* 481–482.
65. Zimmerman, H. E.; Grunewald, G. L. *J. Am. Chem. Soc.* **1966,** *88,* 183–184.
66. Fukui, K.; Yonezawa, T.; Shingu, H. *J. Chem. Phys.* **1952,** *20,* 722–725.
67. Fukui, K.; Nagata, C.; Yonezawa, T.; Kato, H.; Morokuma, K. *J. Chem. Phys.* **1959,** *31,* 287–293.
68. *Frontier Orbitals and Reaction Paths*; Fukui, K.; Fujimoto, H., Eds.; World Scientific Series in 20th Century Chemistry; World Scientific: Singapore, 1997; Vol. 7, pp. 25–214.
69. Woodward, R. B.; Hoffmann, R. *J. Am. Chem. Soc.* **1965,** *87,* 395-397.
70. Woodward, R. B.; Hoffmann, R. *The Conservation of Orbital Symmetry*; Verlag Chemie GmbH: Weinhein, 1970.
71. Hoffmann, R. *Angew. Chem. Int. Ed. Engl.* **2004,** *43,* 6586–6590.
72. Nobel Prize in Chemistry 1981: to Kenichi Fukui and Roald Hoffman "for their theories … concerning the course of chemical reactions."
73. Vogel, E.; Grimme, W; Dinné, E. *Tetrahedron Lett.* **1965,** *7,* 391–395.
74. First Nobel Prize in Chemistry 1901: to Jacobus Henricus van't Hoff "in recognition of…the discovery of the laws of chemical dynamics and osmotic pressure in solutions."
75. Van't Hoff, J. H. *Bull. Soc. Chim. Fr.* **1875,** *23,* 295–301.
76. Baeyer, A. *Ber. Dtsch. Chem. Ges.* **1885,** *18,* 2269–2285.
77. Nobel Prize in Chemistry 1961: to Derek H. R. Barton and Odd Hassel "for their contributions to the development of the concept of conformation."
78. Barton, D. H. R. *Experientia* **1950,** *6,* 316–320.
79. Deslongchamps, P. *Stereoelectronic effects in organic chemistry;* Organic Chemistry Series; Pergamon Press: Oxford, UK, 1983; Vol. 1.
80. Johnson, F. *Chem. Rev.* **1968,** *68,* 375–413.
81. Hoffmann, R. W. *Chem. Rev.* **1989,** *89,* 1841–1860.
82. Merrifield, B. *J. Am. Chem. Soc.* **1963,** *85,* 2149–2154.
83. Corey, E. J.; Venkateswarlu, A. *J. Am. Chem. Soc.* **1972,** *94,* 6190–6191.
84. Huang-Minlon *J. Am. Chem. Soc.* **1946,** *68,* 2487–2488.
85. Lloyd, N. C.; Morgan, H. W.; Nicholson, B. K.; Ronimus, R. S. *Angew. Chem. Int. Ed. Engl.* **2005,** *44,* 941–944.
86. Holton, R. A. U.S. Patent 5,175,315, 1992.
87. Second Nobel Prize in Chemistry 1902: to Hermann Emil Fischer "in recognition of…his work on sugar and purine syntheses."
88. Fischer, E. *Ber. Dtsch. Chem. Ges.* **1890,** *23,* 799–805.
89. Komppa, G. *Ber. Dtsch. Chem. Ges.* **1903,** *36,* 4332–4335.
90. Perkin, W. H.; Thorpe, J. F. *J. Chem. Soc.* **1904,** 128 –148.
91. Perkin, W. H. *J. Chem. Soc.* **1904,** 654–671.

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

92. Nobel Prize in Chemistry 1947: to Sir Robert Robinson "for his investigations on plant products of biological importance, especially the alkaloids."
93. Robinson, R. *J. Chem. Soc.* **1917,** 762–768.
94. Nobel Prize in Chemistry 1930: to Hans Fischer "for his researches into the constitution of haemin and chlorophyll and especially for his synthesis of haemin."
95. Fischer, H.; Kirstahler, A. *Justus Liebigs Ann. Chem.* **1928,** *466,* 178.
96. Bachmann, W. E.; Cole, W.; War Wilds, A. L. *J. Am. Chem. Soc.* **1939,** *61,* 974–975.
97. Rabe, P. *Ber. Dtsch. Chem. Ges.* **1908,** *41,* 62–70.
98. Nobel Prize in Chemistry 1965: to Robert B. Woodward "for his outstanding achievements in the art of organic synthesis."
99. Woodward, R. B.; Doering, W. E. *J. Am. Chem. Soc.* **1945,** *67,* 860–874.
100. Seeman, J. I. . *Angew. Chem. Int. Ed. Engl.* **2007,** *46,* 1378–1413.
101. Smith A. C.; Williams, R. M. *Angew. Chem. Int. Ed. Engl.* **2008,** *47,* 1736–1740.
102. Woodward, R. B.; Sondheimer, F.; Taub, D. *J. Am. Chem. Soc.* **1951,** *73,* 4057.
103. Gates, M.; Tschudi, G. *J. Am. Chem. Soc.* **1952,** *74,* 1109–1110.
104. Stork, G.; Clarke, F. H. *J. Am. Chem. Soc.* **1955,** *77,* 1072–1073.
105. Woodward, R. B.; Bader, F. E.; Bickel, H.; Frey, A. J.; Kierstead, R. W. *J. Am. Chem. Soc.* **1956,** *78,* 2023–2025.
106. Sheehan, J. C.; Henery-Logan, K. R. *J. Am. Chem. Soc.* **1957,** *79,* 1262–1263.
107. Woodward, R. B. *Pure Appl. Chem.* **1968,** *17,* 519–547.
108. Woodward, R. B. *Pure Appl. Chem.* **1971,** *25,* 283–304.
109. Woodward, R. B. *Pure Appl. Chem.* **1973,** *33,* 145–177.
110. Eschenmoser, A.; Wintner, C. E. *Science* **1977,** *196,* 1410–1420.
111. Corey, E. J.; Trybulski, E. J.; Melvin, L. S., Jr.; Nicolaou, K. C.; Secrist, J. A.; Lett, R.; Sheldrake, P. W.; Falck, J. R.; Brunelle, D. J.; Haslanger, M. F.; Kim, S.; Yoo, S. *J. Am. Chem. Soc.* **1978,** *100,* 4618–4620.
112. Corey, E. J.; Kim, S.; Yoo, S.; Nicolaou, K. C.; Melvin, L. S., Jr.; Brunelle, D. J.; Falck, J. R.; Trybulski, E. J.; Lett, R.; Sheldrake, P. W. *J. Am. Chem. Soc.* **1978,** *100,* 4620–4622.
113. Paquette, L. A.; Ternansky, R. J.; Balogh, D. W. *J. Am. Chem. Soc.* **1982,** *104,* 4502–4503.
114. Paquette, L. A.; Ternansky, R. J.; Balogh, D. W.; Kentgen, G. *J. Am. Chem. Soc.* **1983,** *105,* 5446–5450.
115. Paquette, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982,** *79,* 4495–4500.
116. Armstrong, R. W.; Beau, J.-M.; Cheon, S. H.; Christ, W. J.; Fujioka, H.; Ham, W.-H.; Hawkins, L. D.; Jin, H.; Kang, S. H.; Kishi, Y.; Martinelli, M. J.; McWhorter, W. W., Jr.; Mizuno, M.; Nakata, M.; Stutz, A. E.; Talamas, F. X.; Taniguchi, M.; Tino, J. A.; Ueda, K.; Uenishi, J.; White, J. B.; Yonaga, M. *J. Am. Chem. Soc.* **1989,** *111,* 7525–7530.
117. Armstrong, R. W.; Beau, J.-M.; Cheon, S. H.; Christ, W. J.; Fujioka, H.; Ham, W.-H.; Hawkins, L. D.; Jin, H.; Kang, S. H.; Kishi, Y.; Martinelli,

M. J.; Mc Whorter, W. W., Jr.; Mizuno, M.; Nakata, M.; Stutz, A. E.; Talamas, F. X.; Taniguchi, M.; Tino, J. A.; Ueda, K.; Uenishi, J.; White, J. B.; Yonaga, M. *J. Am. Chem. Soc.* **1989,** *111,* 7530–7533.
118. Suh, E. M.; Kishi, Y. *J. Am. Chem. Soc.* **1994,** *116,* 11205–11206.
119. Kishi, Y. *Pure Appl. Chem.* **1989,** *61,* 313–324.
120. La Clair, J. J. *Angew. Chem. Int. Ed. Engl.* **2006,** *45,* 2769–2773.
121. Rychnovsky, S. D. *Org. Lett.* **2006,** *8,* 2895–2898.
122. Porco, J. A. Jr.; Su, S.; Lei, X.; Bardhan, S.; Rychnovsky, S. D. *Angew. Chem. Int. Ed. Engl.* **2006,** *45,* 5790–5792.
123. Halford, B. *Chem. & Engin. News* **2006,** *84(31),* p.11.
124. Youngsaye, W.; Lowe, J. T.; Pohlki, F.; Ralifo, P.; Panek, J. S. *Angew. Chem. Int. Ed. Engl.* **2007,** *46,* 9211–9214.
125. Custar, D. W.; Zabawa, T. P.; Scheidt, K. A. *J. Am. Chem. Soc.* **2008,** *130,* 804–805.
126. Jiang, X.; Liu, B.; Lebreton, S.; De Brabander, J. K. *J. Am. Chem. Soc.* **2007,** *129,* 6386–6387.
127. Nicolaou, K. C.; Guduru, R.; Sun, Y.-P.; Banerji, B.; Chen, D. Y.-K. *Angew. Chem. Int. Ed. Engl.* **2007,** *46,* 5896–5900.
128. Johnson, W. S. *Acc. Chem. Res.* **1968,** *1,* 1–8.
129. Heathcock, C. H.; Hansen, M. M.; Ruggeri, R. B.; Kath, J. C. *J. Org. Chem.* **1992,** *57,* 2544–2553.
130. Stork, G.; Burgstahler, A. W. *J. Am. Chem. Soc.* **1955,** *77,* 5068–5077.
131. Eschenmoser, A.; Ruzicka, L.; Jeger, O.; Arigoni, D. *Helv. Chim. Acta.* **1955,** *38,* 1890–1904.
132. Stadler, P. A.; Eschenmoser, A.; Schniz, H.; Stork, G. *Helv. Chim. Acta.* **1957,** *40,* 2191.
133. Anastas, P. T.; Warner, J. C. *Green Chemistry: Theory and Practice;* Oxford University Press: New York, 2000.
134. Trost, B. M. *Angew. Chem. Int. Ed. Engl.* **1995,** *39,* 259–281.
135. Nicolaou, K. C.; Sorensen, E. J. *Classics in Total Synthesis;* VCH: Weinheim, Germany, 1996.
136. Trost, B. M. *Science* **1985,** *227,* 908–916.

Downloaded by STANFORD UNIV GREEN LIBR on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch010

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

Chapter 11

The Role of Evolution in the Discovery of New Drugs and Chemicals

S. R. Rajski[1] and Ben Shen[1,2,*]

[1]Division of Pharmaceutical Sciences, University of Wisconsin National Cooperative Drug Discovery Group, School of Pharmacy, University of Wisconsin-Madison, 777 Highland Ave. , Madison, Wisconsin 53705-2222.
[2]Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin, 53706-1322.

Evolution favors organisms capable of generating and retaining chemical diversity at low cellular cost. We present here three examples of diversity-oriented biosynthetic pathways that, taken together, represent a snapshot of biosynthetic evolution; (i) the evolving mechanisms and structures of polyketide sythetases (PKS) as reflected by the enediyne PKSE; (ii) the predicted versus actual roles of SgcD/SgcG and SgcC4 in production of the enediyne natural product C-1027; and (iii) the production and bioactivity of the C-1027 analogs produced with knowledge of the C-1027 biosynthetic machinery.

What does the study and exploitation of natural products biosynthesis tell us about evolution? Moreover, how is it that evolutionary concepts and applications now manifest themselves in the arena of biosynthetic chemistry? Not surprisingly, advances in microbiology and molecular biology, bioinformatics, genetics, analytical methods and highthroughput technologies have profoundly advanced the field of natural products biosynthesis. The enhanced rate at which discoveries in biosynthesis are now made illustrate more clearly than ever how biosynthetic machineries are evolutionarily linked or divergent and also how the field of natural products biosynthesis is evolving.

That genes coding for biosynthesis of natural products in bacteria are, more often than not, clustered within the same genomic space hastens the identification of complete gene clusters but does not, in and of itself, speak to the advantage of studying biosynthetic pathways from an evolutionary perspective (*1*). Rather, it is the speed with which they evolve due to the short replication times of their microbial hosts and propensity for horizontal transfer between hosts that makes biosynthetic gene clusters such ideal systems to study evolutionarily (*1,2*). Moreover, biosynthetic clusters produce quantifiable phenotypic readouts in the form of small, rigorously characterizable molecules. Biosynthetic pathways for many small molecules are understood in sufficient detail such that the role of each gene in producing a given molecule can usually be assigned; this level of understanding is attributable to the evolution of bioinformatics. More telling from the standpoint of enzyme evolution and by inference, natural products evolution, are those instances in which enzymatic activities differ from those predicted on the basis of bioinformatics. Small, but easily detectable differences among natural products can thus serve as a beacon for discovery of new enzymatic activities that translate directly into an enhanced appreciation for enzyme evolution.

The evolution of secondary metabolism has been postulated to favor organisms that generate and retain chemical diversity at a low cost to the cell's metabolic machinery and without posing a threat of toxicity (*3,4*). In line with this rationale, organisms that make and "screen" a large number of metabolites will have an increased likelihood of enhanced fitness simply because the greater the chemical diversity, the greater the chances of producing the rare chemical with a useful, potent biological activity (*3,4*). There is, presumably, no reason that host toxins are produced, or if they are, the host organism co-evolves in parallel, resistance mechanisms to prevent deleterious effects. Originally coined the "screening" model of secondary metabolite biosynthesis, these tenets may best be described as a "diversity-based" module to emphasize the nature of the biosynthetic pathways rather than the way in which their products are used. Several features of natural product biosynthesis support the diversity-based model, especially the large number of natural products with no known activity (or at least no known potent activity), the tendency of natural product pathways to produce a suite of molecules, and the widespread use of branched and matrix biosynthetic pathways to share metabolic and genetic costs. Exemplary in this respect is the biosynthesis of the enediyne family of antibiotics some of which are known to display potent bioactivity.

The enediynes are characterized structurally by an enediyne core unit possessing two acetylenic groups conjugated to a double bond or incipient

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

double bond within a 9- or 10-membered ring (Figure 1). To date, at least six 10-membered enediynes (Figure 1a) and five different 9-membered enediynes (Figure 1b) have been characterized (*5-7*). In general, these substances contain three distinct structural elements: a DNA recognition unit serving to deliver the metabolite to its target DNA; an activation motif, which, upon triggering, induces cycloaromatization of the enediyne moiety; and the enediyne itself which cycloaromatizes to a highly reactive diradical species and, in the presence of DNA, results in oxidative strand scission via H-atom abstraction from the deoxyribose backbone (*6,8*).

Cycloaromatization is most often triggered either in a nucleophilic manifold (*9,10*) as is known for neocarzinostatin (Figure 2a) or via a strain-induced motif wherein relaxation of molecular strain is afforded by enediyne cyclization (*5,11,12*) (Figure 2b). Reaction of diradicals resulting from either activation motif with the sugar phosphate backbone of DNA affords carbon-centered radicals, which upon reaction with molecular O_2, produces highly labile lesions that segue into double stranded breaks (DSBs) (Figure 2). In addition to the structurally characterized enediynes, the enediyne scaffold has been proposed to play a role in the biosynthesis of the sporolides and cyanosporasides (Figure 1c) on the basis of biosynthetic gene cluster characterization and specific similarities to gene clusters associated with bona fide enediyne natural products (*13,14*).

A member of the 9-membered enediyne class, C-1027 is one of the most cytotoxic natural products ever discovered and, like its relatives neocarzinostatin, maduropeptin, and kedarcidin, is isolated as a chromoprotein complex (*5,6*). Isolated from *Streptomyces globisporus* as a noncovalent apoprotein (CagA)-chromophore complex, the C-1027 chromophore consists of an enediyne core, a deoxy amino sugar, a β-amino acid, (circled in Figure 1b) and a benzoxazolinate moiety (boxed in Figure 1b). The enediyne core induces double-stranded DNA breaks upon cycloaromatization while the other moieties are critical in enhancing bioactivity. This is most clearly the case with the benzoxazolinate moiety, which specifically binds to CagA, hence providing stability to the enediyne core, and is essential for both binding and intercalating the minor groove of DNA. Our efforts to understand C-1027 biosynthesis continue to shed light on the impact of evolution upon drug discovery and the unveiling of new chemical entities from natural sources. Herein, we discuss three topics related to C-1027 that illuminate the role of evolution in this compound's biosynthesis and development. Specifically, (i) the evolving mechanisms and structures of polyketide sythases (PKS) as reflected by the enediyne PKSE, (ii) the predicted versus actual roles of SgcD/SgcG and SgcC4 in production of the C-1027 benzoxazolinate and β-amino acid components, respectively, and (iii) the production and bioactivity of non-natural C-1027 analogs produced with knowledge of the C-1027 biosynthetic machinery are highlighted.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

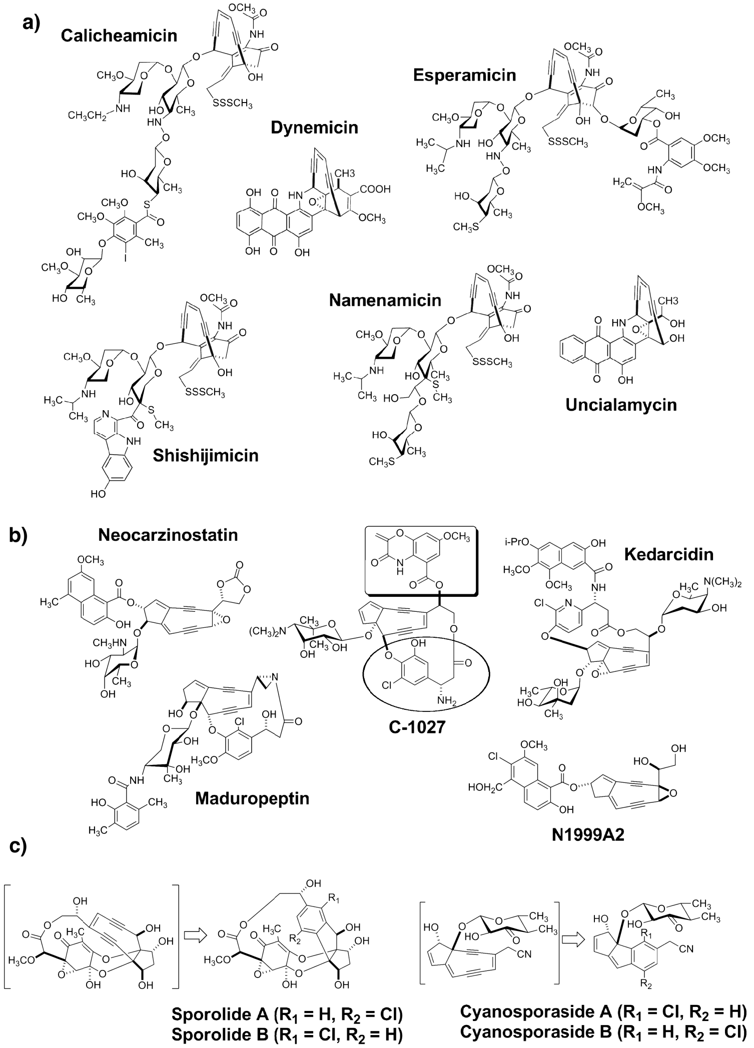

a)
Calicheamicin
Dynemicin
Esperamicin
Namenamicin
Uncialamycin
Shishijimicin
b)
Neocarzinostatin
Kedarcidin
C-1027
Maduropeptin
N1999A2
c)
Sporolide A (R_1 = H, R_2 = Cl)
Sporolide B (R_1 = Cl, R_2 = H)
Cyanosporaside A (R_1 = Cl, R_2 = H)
Cyanosporaside B (R_1 = H, R_2 = Cl)

Figure 1. Naturally occurring enediynes and natural products for which enediynes have been proposed as biosynthetic intermediates. (a) The 10-membered family includes calicheamicin, dynemicin, esperamicin, shishijimicin, namenamicin, and uncialamycin. (b) The 9-membered enediynes include neocarzinostatin (NCS), C-1027, maduropeptin, kedarcidin, and N1999A2. (c) Sporolides A and B and cyanosporasides A and B are all hypothesized to originate from enediyne precursors (bracketed) based on the presence of the enediyne polyketide synthase (PKSE) in the Salinaspora genome of each producing organism.

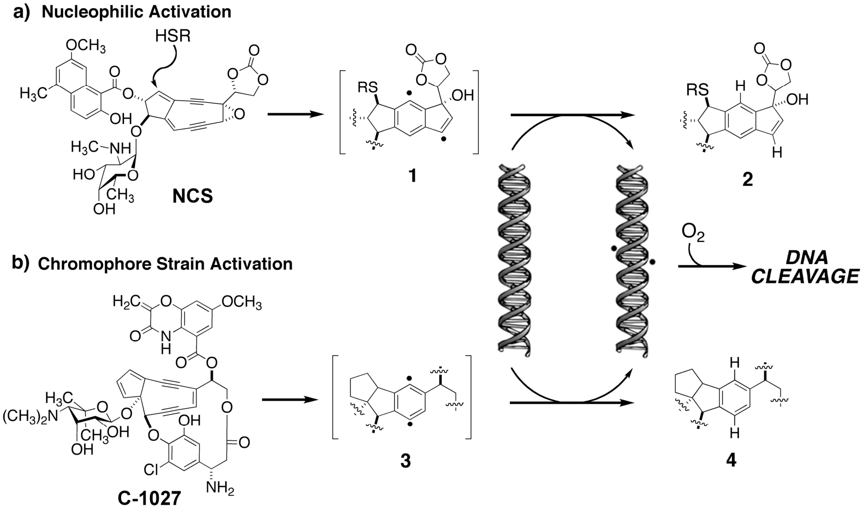


Figure 2. Depiction of O_2-dependent DNA strand cleavage induced by activated enediynes. (a) Mechanism of nucleophilic activation and subsequent radical chemistry leading to DNA strand breaks for NCS. (b) Mechanism of strain relaxation-triggered activation and subsequent radical chemistry leading to DNA strand breaks for C-1027.

The Enediyne Polyketide Synthase (PKSE): Evolutionary Origins and a Means by Which to Find New Bioactive Enediynes

Feeding experiments with ^{13}C-labeled precursors established many years ago that both the nine- and ten-membered enediyne core structures are derived from eight head-to-tail acetate units (*6*). However, the precise construction of the enediyne cores using acetate remained a topic for debate as recently as six years ago. Whereas such core structures might be envisioned to result from de novo polyketide biosynthesis, the possibility also existed that both the nine- and ten-membered cores might result from degradation of fatty acid precursors (*6*). The latter hypothesis was fueled in large part by the fact that enediynes bear no structural resemblance to any other characterized polyketides. The phylogenetic isolation of the enediyne core biosynthetic enzymes from other known polyketide synthases, but potential relatedness between those enzymes responsible for nine- and ten-membered enediynes suggests that the enzymes driving enediyne biosynthesis represent unique evolutionary branch points for a larger class of enzymes that either have lost the ability to produce enediynes or that produce enediynes not yet discovered. On this and many other counts, the concept of evolution explains and motivates studies in enediyne biosynthesis.

Important insight into the biosynthetic origins for both classes of enediyne-containing natural products was obtained upon cloning of the C-1027 biosynthetic gene cluster (*15*). Sequencing of the cluster revealed only one PKS gene, *sgcE*, whose deduced product was found to consist of five domains, four

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

of which [ketosynthase (KS), acyltransferase (AT), ketoreductase (KR), and dehydratase (DH)] are characteristic of previously deduced type I PKSs. The fifth domain, residing at the C-terminus (initially called the terminal domain, TD) was subsequently found to be unique to enediyne PKSs. Speculations at the time of cluster elucidation indicated that the region between the AT and KR domains might contain an acyl carrier protein (ACP) domain and that the TD might serve the role of a 4'-phosphopantetheinyl transferase (PPT). We have recently confirmed experimentally that the TD is, in fact, a PPT domain (*16*). Notably, polyunsaturated fatty acid synthases are known to contain between 5 and 9 ACP domains in the region between their AT and KR domains (*17*). SgcE involvement in the biosynthesis of C-1027 was confirmed by gene inactivation and complementation: inactivation of the *sgcE* gene ablated C-1027 production in *S. globisporus* whereas its overexpression in *ΔsgcE* restored production of the natural product (*5,15*). We envisioned SgcE to catalyze the assembly of a linear polyunsaturated intermediate from acyl-CoA precursors in an iterative process. The nascent intermediate could then be subsequently desaturated (by other enzyme activities) to furnish the two alkyne groups and cyclized to afford the enediyne core. The latter postulate was supported by the finding that a group of five to ten genes, flanking the *sgcE* enediyne PKS gene, is highly conserved among all enediyne gene clusters characterized at the time (*18*). These genes were found to encode assorted oxidoreductases or proteins of unknown functions that are associated only with enediyne biosynthesis, serving as candidates for processing the nascent linear polyketide intermediate into a cyclic enediyne intermediate. Significantly, the chromosomal arrangement of these genes across the six characterized biosynthetic clusters known at the time revealed a high degree of similarity from one enediyne producer to the next despite the lack of absolute certainty regarding the function of each discrete gene (*18*). The chromosomal arrangements determined for the C-1027 and NCS biosynthetic clusters are shown in Figure 3a.

Coinciding with our efforts to elucidate SgcE from the C-1027 cluster, Thorson and co-workers cloned the biosynthetic gene cluster from *Micromonospora calichensis* for the 10-membered enediyne calicheamicin (*19*). Two PKS genes, *calE8* and *calO5*, were identified within the calicheamicin cluster. The *calO5* gene was found to encode for a synthase associated with orsellinic acid production and was thus negated from involvement in enediyne synthesis. However, *calE8* involvement in calicheamicin biosynthesis was definitively established by gene inactivation experiments. Remarkably, CalE8 was found to exhibit head-to-tail sequence homology to SgcE and also possessed the same domain organization of KS, AT, KR, DH and PPT observed in SgcE (Figure 3b) (*6,18*).

The structural and functional relatedness of SgcE and CalE8 suggested a common polyketide pathway for the biosynthesis of both nine and ten-membered enediynes, despite the fact that their incorporation patterns by ^{13}C-labeled acetate feeding experiments were distinct. The two carbons that make up each triple bond of neocarzinostatin and C-1027 (both nine-membered enediynes) are derived from intact acetate units, whereas those of calicheamicin are known to be derived from adjacent acetate units. Like SgcE, the CalE8 PKS could be envisioned to produce a nascent polyunsaturated intermediate from

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

acyl CoA precursors via an iterative process as indicated for production of polyketide **5** (Figure 3b) (*6*). Indeeed, on the basis of findings for both the C-1027 and calicheamicin systems the general paradigm shown in Figure 3 could be envisioned for all enediyne producers. Production of general intermediate **5** by a minimal PKSE unit would provide to accessory proteins unique to either 9- or 10-membered enediyne producing organisms an advanced intermediate by which either class of compounds could be made through the agency of **6** (for 9-membered compounds) or **7** and/or **8** (for 10-membered enediynes). It thus was recognized that SgcE and CalE8, represent a novel family of iterative type I PKSs, establishing a new paradigm for enediyne biosynthesis.

The similarities observed for SgcE and CalE8 inspired significantly greater effort in the search for enediyne biosynthetic gene clusters. High-throughput genome-scanning methods have been used to detect and analyze gene clusters involved in natural product biosynthesis, and in fact, multiple genetic loci homologous to those of *sgcE* and *calE8*, from organisms not previously recognized as enediyne producers were identified (*18*). A set of 12 functional enediyne PKS genes were found dispersed among genomes of ~50 widely diverse organisms. NcsE, SgcE, MadE (unshaded circles, Figure 4), CalE8, EspE and DynE (unshaded boxes, Figure 4) were known prior to genome scanning of the actinomycete library. This knowledge enabled phylogenetic analysis of the entire set of cloned bona fide enediyne PKSs from which was unveiled the separation of this set into two main branches. Segregation of the branch coincides with the distinction of the core enediyne ring size (Figure 3b) (*6*). The enediyne PKSs therefore must have evolved from the same ancestor, divergent evolution from which affords the 9- and 10-member-specific accessory enzymes. This divergent relationship provides a basis for a predictive model from which the fundamental structure of unknown enediyne cores can be predicted directly based on a cloned enediyne *pksE* gene, which in turn can be accessed rapidly via PCR. That only the 9-membered ring enediynes are associated with apoproteins (with the exception of N1999A2) also suggests a unique evolutionary branch point in enediyne production.

The familial model (Figure 4) resulting from phylogenetic analysis suggested that five of the unknown potential enediyne PKSs were expected to produce 9-membered products (*macE*, *ghaE*, *strE*, *megE*, and *ecoE*), whereas the remaining seven are genotypically predictive of 10-membered core formation (*micE*, *kitE*, *cavE*, *aerE*, *kanE*, *cirE*, and *oriE*) (*6*). Subsequent fermentation efforts with a subset of eight enediyne producing candidate organisms (PKSEs shaded, Figure 4) followed by biochemical induction assays (BIAs) routinely used to assess production of DNA-damaging small molecules revealed that three out of the eight microorganisms housing the genetic machinery for PKSEs shown in Figure 4 are very likely enediyne producers (*18*). Enediyne production in *Amycolatopsis orientalis (housing oriE)*, *Streptomyces ghanaensis (housing ghaE)*, and *Streptomyces citricolor (housing cirE)* has been found to be inducible under optimized growth conditions (*18*). Additionally, two new actinomycetes in the original panel of 50 not highlighted in Figure 4 were implicated as likely enediyne producers on the basis of BIAs. The success of this study illustrates the exceptional speed and potential power of

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

enediyne-producer familial classification and highlights a new direction in which the practices of drug discovery are evolving.

The realization that *sgcE* and *calE8* are not isolated examples of the evolution of enediyne biosynthesis was accompanied by our development and application of the PCR-based approach to access the PKSEs and its associated accessory genes directly. For example, the enzymes MadE, DynE and EspE involved in the biosynthesis of maduropeptin, dynemicin and esperamicin respectively, have been found to share the common enediyne PKSE domain organization (KS, AT, ACP, KR, DH, and PPT) first observed with SgcE from the C-1027 cluster (*6,18-20*). Indeed, all PKSEs characterized to date display this domain organization. However, only recently has the functional assignment of the PPT and ACP domains as originally hypothesized for PKSE been experimentally supported (*16*).

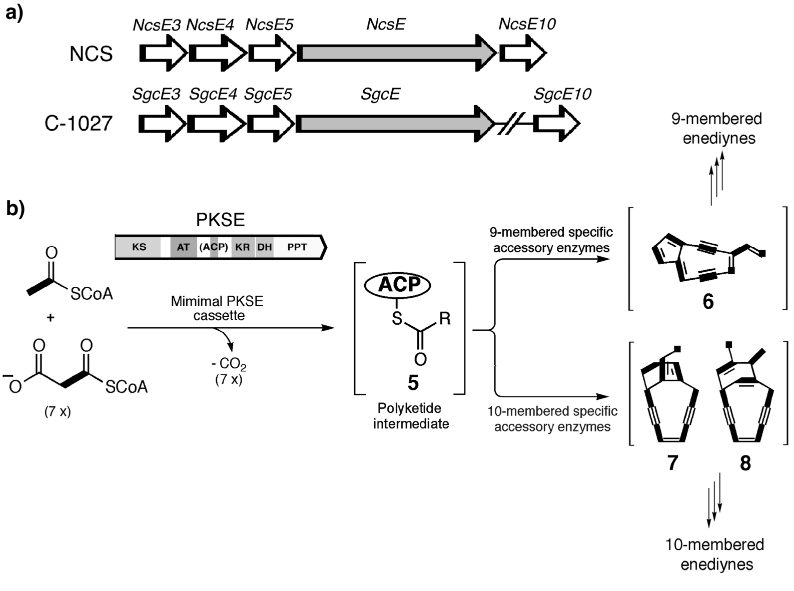

a)
NcsE3 NcsE4 NcsE5
NcsE
NcsE10
NCS
SgcE3 SgcE4 SgcE5
SgcE
SgcE10
C-1027
9-membered enediynes
b)
PKSE
KS
AT
(ACP)
KR
DH
PPT
O
SCoA
+
O
O
O
SCoA
(7 x)
Mimimal PKSE cassette
- CO2
(7 x)
ACP
S
R
O
5
Polyketide intermediate
9-membered specific accessory enzymes
10-membered specific accessory enzymes
6
7
8
10-membered enediynes

*Figure 3. Minimal PKSE cassettes for the C-1027 and NCS biosynthetic loci and a general paradigm for enediyne biosynthesis. (a) Organization of the enediyne gene cassettes found in the neocarzinostatin (NCS), and C-1027 loci. (b) Unified scheme for enediyne core biosynthesis carried out by the PKSE and associated accessory enzymes responsible for diverging common intermediate 5 into either 9- or 10-membered enediyne core pathways through the agency of carbacycles **6**, 7 or **8**. PKSE domain layout indicated is highly conserved for all PKSEs characterized to date.*

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

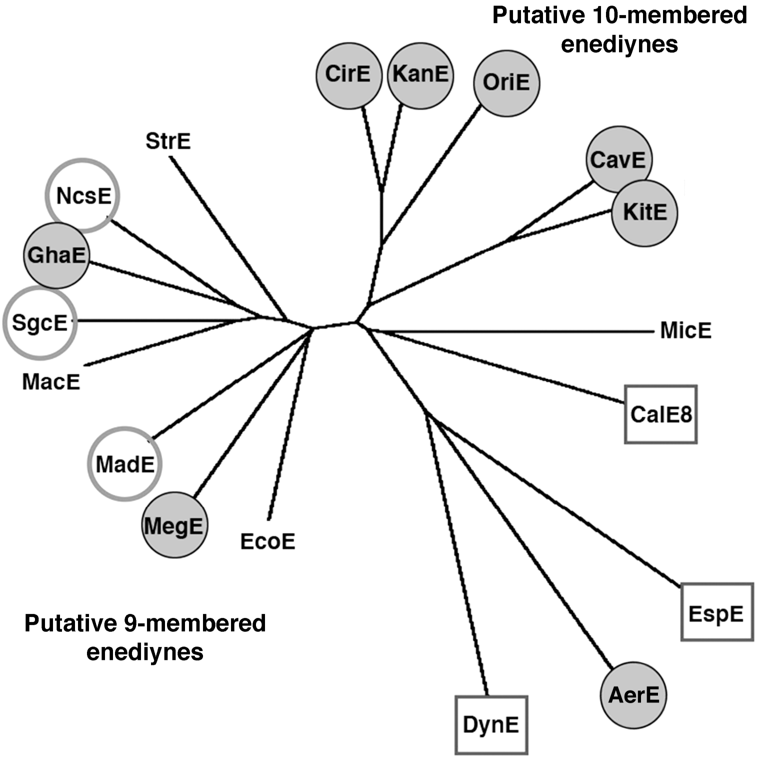


Figure 4. Enediyne producer familial model based on phylogenetic comparison of bona fide enediyne minimal PKSs. Organisms possessing PKSEs in shaded circles were subjected to fermentation followed by BIA to investigate for putative enediyne production. PKSEs indicated with unshaded circles or boxes correlate to well established enediyne producing microbes. Given in parentheses are GenBank accession numbers: Lechevalieria aerocolonigenes AerE (AAO25864); Micromonospora echinospora CalE8 (AAM94794); Streptomyces cavourensis subsp. washingtonensis CavE (AAO25858); Streptomyces citricolor CirE (AAO25874); Micromonospora chersina DynE (AAN79725); Streptomyces sp. 100/Eco52 EcoE (AAO25879); Actinomadura verrucosospora EspE (AY267372); Streptomyces ghanaensis GhaE (AAO25844); Streptomyces kaniharaensis KanE (AAO25869); Kitasatospora sp. CECT 4991 KitE (AAO25848); Streptomyces macromomyceticus MacE (AAO25894); Actinomadura madurae MadE (AY271660); Micromonospora megalomicea subsp. nigra MegE (AAO25852); Micromonospora sp. 046/Eco11 MicE (AAO25884); Streptomyces carzinostaticus NcsE (AAM78012); Amycolatopsis orientalis OriE (AAO25836); Streptomyces globisporus SgcE (AY048920); Streptomyces sp. 171/Eco105 StrE (AAO25889).

The first biosynthetic step leading to the enediyne core is speculated to be catalyzed by a new type of iterative PKSE characteristic of the enediyne family as suggested by the familial model resulting from phylogenetic analyses. PKSs catalyze decarboxylative condensation in a fashion similar to that of fatty acid synthases (FASs). Both FASs and PKSs thus share homologous KS domains responsible for decarboxylation and condensation, and (AT) domains, responsible for selection and loading of an acyl CoA (CoA)-substrate to a free thiol of the ACP. The ACP thiol is generated posttranslationally by covalent tethering of the 4'-phosphopantetheine (4'-PP) component of CoA onto a conserved serine (Ser), and this reaction, catalyzed by all PPTs, is essential for PKS and FAS systems with the exception of type III PKSs (*21*). Unlike FASs, which contain a full set of KR, DH, and enoylreductase (ER) domains to reduce the resulting carbonyl to saturation, PKSs can have from all to none of these activities, resulting in a product with various levels of oxidation (*16*).

Domain similarities between FASs and PKSs have long fueled the idea that enediynes might originate largely through the actions of FASs. Thus, unique characteristics of PKSEs have played a large role in accurately defining the origins of enediyne natural products. As a result, we have very recently dissected the structure and activities of the suspected ACP and PPT domains of the PKSEs responsible for both C-1027 and NCS biosynthesis (*16*). The careful application of bioinformatics, gene inactivation and cross-complementation, and high resolution mass spectrometry have allowed a number of significant conclusions to be made regarding domain organization and activity of SgcE and NcsE. First, it was determined that the domains between the AT and KR domains for both SgcE and NcsE are bona fide ACPs supporting early assertions that the initial steps of enediyne core biosynthesis are catalyzed by PKSE that is ACP-dependent; S974 and S1007 are the active site serines subjected to phosphopantetheinylation necessary for C-1027 and NCS biosynthesis respectively (Figure 5). Secondly, the function of the C-terminal domain previously denoted as TD was determined to be that of a PPTase responsible for attachment of the 4'-PP prosthetic group from CoA to the ACP (Figure 5); this was found to be the case for both C-1027 and NCS producing PKSs and is very likely the case for most PKSEs. This, in particular, sets the PKSEs apart from all known PKSs and FASs since no other examples of "self-phosphopantetheinylating" PKSs or FASs are known. Thirdly, the nine-membered enediyne PKSEs from both C-1027 and NCS clusters were found, in non-native producing organisms, to catalyze iterative decarboxylative condensation, ketoreduction, and dehydrations to afford an ACP-bound polyene intermediate that could be hydrolyzed by cognate thioesterase (TE) domains (designated either SgcE10 or NcsE10) to yield the polyene 1,3,5,7,9,11,13-pentadecaheptaene, a logical building block for both C-1027 and NCS (Figure 5). Finally; it was determined that the PKSE is interchangeable within the nine-membered family of enediynes, suggesting that the biosynthesis of the nine-membered enediyne core occurs through a common polyene intermediate that is subsequently decorated in each case via different post-PKS tailoring enzymes. These findings further confirm the molecular logic of enediyne PKSEs as being distinct from all known PKS and FAS paradigms (*16*) and no doubt shed light upon the evolution of enediyne natural products in addition to supporting the

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

validity of previous phylogenetic analyses that show a clear biosynthetic relatedness among enediyne producers.

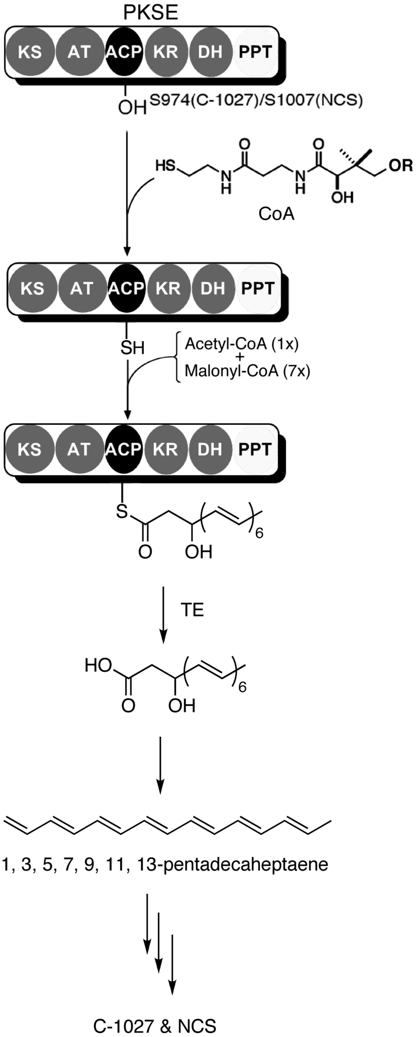

PKSE
KS
AT
ACP
KR
DH
PPT
OH
S974(C-1027)/S1007(NCS)
HS
N
H
O
OH
OR
CoA
SH
Acetyl-CoA (1x)
+
Malonyl-CoA (7x)
S
6
TE
HO
1, 3, 5, 7, 9, 11, 13-pentadecaheptaene
C-1027 & NCS

Figure 5. Final domain assignments of nine-membered enediyne PKSEs from C-1027 producer S. globisporus and the NCS producer S. carzinostaticus. Active site serines within ACP domains are phosphopantetheinylated by PPT domain (in cis) followed by production of the polyene precursor to the enediyne core. R, 3', 5'-adenosine diphosphate.

Recent application of PKSE data sets and the noted PCR-based strategy has revealed that the biosynthetic gene cluster for sporolides A and B contains two loci encoding PKSEs (*13*). Produced by *Salinispora tropica*, sporolides A and B do not contain enediynes per se (Figure 1c). Rather, each substance contains a tricyclic motif that is proposed to originate biosynthetically from a transient

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

enediyne, cycloaromatization of which affords the tricycle (*13*). Although no such enediyne sporolide precursor has yet been identified, the prospect that the PKSEs may be involved in production of natural products devoid of an enediyne moiety broadens the significance, not only of the PKSs themselves, but also their evolution from a common ancestor. Importantly, similar rationale has been invoked for the formation of the tricyclic core of cyanosporasides A and B (*14*) (Figure 1c).

The Relationship of SgcD, SgcG and SgcC4 in C-1027 Production: Emphasis on Enediyne Decorating Moieties Whose Function Diverges with that Predicted by Bioinformatics

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

The benzoxazolinate moiety of C-1027 plays a vital role in both stabilization of the enediyne core via CagA associations as well as in the effective noncovalent association of C-1027 with DNA. The precursor of this moiety was initially proposed to be chorismate based on the uncovering of proteins homologous to anthranilate synthase (AS) components I and II that are located within the C-1027 biosynthetic gene cluster (*22*). ASI, typically annotated as TrpE, catalyzes the first committed step in tryptophan (Trp) biosynthesis: the conversion of chorismate to anthranilate, whereas ASII (TrpG) provides the amino moiety from glutamate (Figure 6). The reaction catalyzed by ASI is the composite of two enzymatic steps: a reversible 1,5-substitution involving amine addition at C2 of chorismate with concomitant loss of the hydroxyl group at C4 to yield 2-amino-2-deoxyisochorismate (ADIC, **10**), followed by irreversible pyruvate elimination to form anthranilate. The overall reaction coordinate proceeds with ADIC as an enzyme bound intermediate, as deduced from (i) the conversion of synthetic ADIC to anthranilate, and (ii) transient accumulation of small amounts of ADIC from a *Salmonella typhimirium* ASI (H398M) mutant (*22*).

In addition to the Trp biosynthetic pathway, chorismate is also the branch point for at least five other pathways found in primary and secondary metabolism. The majority of these pathways are initiated by ASI homologues (Figure 7) (*22*). Two distinct chorismate-using families with no significant sequence homology to ASI are also known; chorismate lyase, the first enzyme in ubiquinone biosynthesis, and chorismate mutase (*16*), the first enzyme in phenylalanine (Phe) and Tyr biosynthesis (*22*). Chorismate is therefore a broadly used biological starting material able to be used not only by multiple enzymes within one class but also by enzymes belonging to different classes.

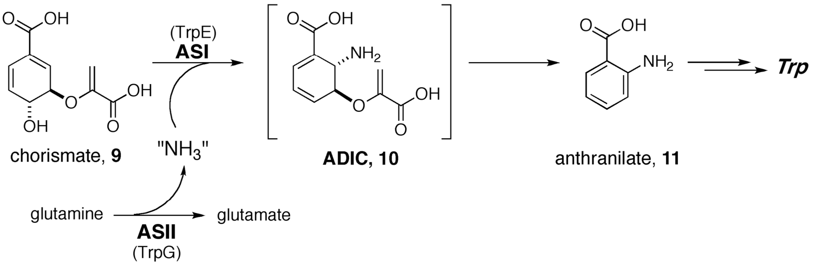

(TrpE)
ASI
"NH3"
glutamine
glutamate
ASII
(TrpG)
chorismate, 9
ADIC, 10
anthranilate, 11
Trp

Figure 6. Biosynthetic pathway for production of tryptophan (TRP) precursor anthranilate from chorismate by anthranilate synthase component I (ASI).

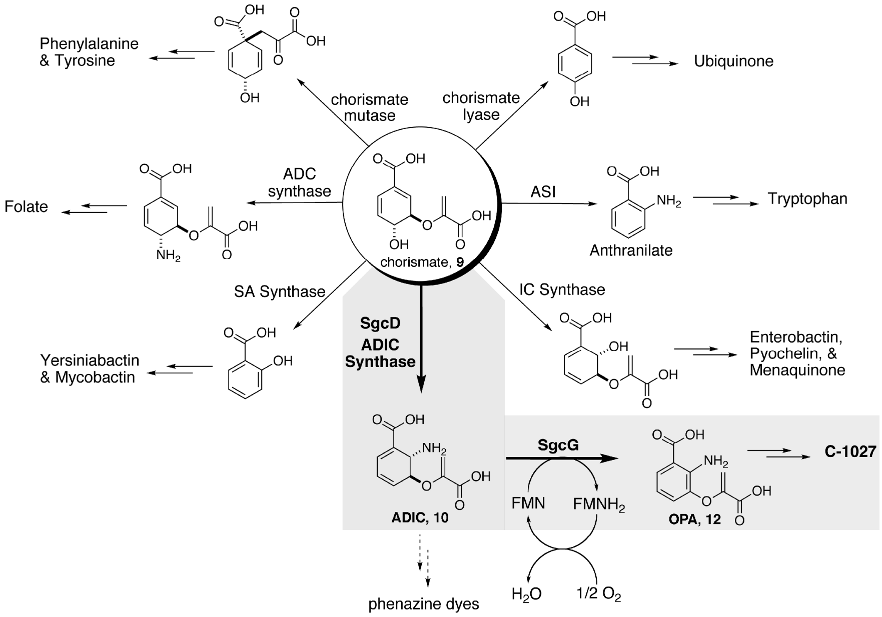

Phenylalanine & Tyrosine
chorismate mutase
chorismate lyase
Ubiquinone
Folate
ADC synthase
ASI
Anthranilate
Tryptophan
chorismate, 9
SA Synthase
IC Synthase
Yersiniabactin & Mycobactin
SgcD ADIC Synthase
Enterobactin, Pyochelin, & Menaquinone
ADIC, 10
SgcG
FMN
FMNH2
OPA, 12
C-1027
phenazine dyes
H2O
1/2 O2

Figure 7. Metabolic pathways that use chorismate as a starting material. The SgcD/SgcG-dependent pathway is specific to C-1027 biosynthesis (shaded) and represents the newest pathway originating from chorismate. ADC, 4-amino-4-deoxychorismate; ASI, anthranilate synthase component I; IC, isochorismate; SA, salicylic acid; ADIC, 2-amino-2-deoxyisochorismate; OPA, 3-enolpyruvylanthranilate.

SgcD and SgcG in Benzoxazolinate Biosynthesis

Bioinformatics revealed that SgcD is a member of the family of chorismate binding enzymes with closest sequence homology to ASI enzymes responsible for the first committed step in Trp biosynthesis (*22*). The *sgcD* gene was found to reside in a putative operon containing five clustered open reading frames (*orf*) previously predicted to be involved in benzoxazolinate biosynthesis

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

(*15*). ASI enzymes from *S. typhimirium*, *E. coli* and *Serratia marcescens* were all found to be closely related homologs to SgcD. Moreover, structural analysis of *S. marcescens* ASI and *S. typhimirium* ASI revealed that active site amino acids involved in binding to benzoate, pyruvate, and Mg^{2+} and a crucial H398 residue are all conserved in SgcD. Along with *sgcD* was also identified a gene termed *sgcG* which bioinformatics revealed to be most similar to the Isf family of iron-sulfur flavoproteins; both genes reside within the same operon and are separated by only one *orf* (*22*). Although the precise function of the Isf family of proteins remains unclear, the closely related WrbA family, also bearing high sequence similarity to SgcG but lacking the iron-sulfur cluster, are well-established NAD(P)H:quinone oxidoreductases. Thus, SgcG, upon initial inspection appeared responsible for executing redox chemistry involved in benzoxazolinate production.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

Given that two available *Streptomyces* genomes are each annotated with multiple, putative ASIs, genetic redundancy within the C-1027 producer *S. globisporus*, we reasoned, could complicate efforts to elucidate SgcD. As such, we opted to directly establish the catalytic function of SgcD by characterizing the enzyme in vitro. Surprisingly, and in contrast to bioinformatic prediction, preliminary experiments using *E. coli* ASI as a control revealed that SgcD was not an ASI. HPLC analysis clearly showed that SgcD converts chorismate to a product with spectral properties identical to ADIC **10** (shaded path, Figure 7), previously identified as a bona fide intermediate en route to ASI by conversions of synthetic ADIC to anthranilate (*22*). This was further substantiated by confirming that SgcD-prepared ADIC can be converted to anthranilate using recombinant *E. coli* ASI. This provided definitive evidence to support SgcD's role as an ADIC synthase despite the initial bioinformatics that pointed to SgcD as an ASI. SgcD's enzymatic activity is not supported by the initial bioinformatics analysis and suggests SgcD as an ASI enzyme whose lyase activity needed for removal of the enolpyruvoyl moiety from chorismate has been evolutionarily lost.

Mutagenesis experiments with *S. typhimirium* ASI and *E. coli* ADC synthase have hinted at the potential existence of a naturally-occurring ADIC synthase. An *S. typhimirium* ASI (H398M) mutant was shown to transiently accumulate ADIC, although this mutant had <1% of the activity of the wild-type enzyme and produced only anthranilate during prolonged incubations (*22*). Interestingly, comparative modeling of SgcD with the *S. typhimirium* ASI structure suggests SgcD is indeed an ASI, since all the active site residues that have been described as essential for ASI catalysis are identical in SgcD, including the histidine (His) residue proposed to participate in pyruvate elimination. A different mutagenesis study using *E. coli* ADC synthase was also successful in generating an artificial ADIC synthase (*23,24*).

An ADIC synthase function has been proposed for PhzE, an ASI homologue required for phenazine biosynthesis (*22*). However, only indirect evidence has been provided for such a naturally-occurring enzyme activity. Such evidence includes findings that (i) ADIC and not anthranilate is converted to phenazine metabolites from cell-free extracts of *E. coli* strains harboring phenazine *orfs*, and that (ii) the activity of recombinant PhzD, catalysis by which is predicted to occur after PhzE, has been examined and functionally assigned as an isochorismatase, (associated with hydrolysis of vinyl ethers) (*22*). It was determined, however, that the PhzD enzyme had nearly identical second order rate constants with respect to ADIC and isochorismate. Further contradictory evidence to assign PhzE as an ADIC synthase has also been reported, including the ability of PhzE to complement an *E. coli* ASI mutant strain and the interchangeability of the *Pseudomonas aeruginosa* ASI homologues—one involved in the biosynthesis of Trp, the other in the phenazine pyocyanin, thus suggesting PhzE is an anthranilate synthase (*22*). While our results have unambiguously established the activity of SgcD as an ADIC synthase, the fact that PhzE and SgcD have no significant sequence homology (<12% identity), along with the aforementioned inconclusive reports regarding phenazine biosynthesis, clearly prevent a functional assignment of PhzE with the currently available data. Conversely, SgcD, despite its predicted function based on bioinformatics as an ASI, is a bona fide ADIC synthase that constitutes yet another biosynthetic use for chorismate.

The discovery of ADIC synthase activity of SgcD suggested the relevance of an oxidoreductase to furnish the aromatic ring of the C-1027 benzoxazolinate moiety. Just downstream of *sgcD* is *sgcG* coding for a predicted [Fe-S], flavin-dependent oxidoreductase. Although the closest homologues of SgcG (the Isf protein family) do not have a defined catalytic function (*22*), this gene product represented the most logical candidate to perform redox chemistry needed to convert ADIC to a more advanced benzoxazolinate precursor. HPLC was used to assess FMN-dependent conversion of purified ADIC by SgcG. As anticipated, the loss of ADIC was accompanied by formation of a new peak with retention times and UV-vis spectra very similar, surprisingly, to anthranilate. That SgcG does not produce anthranilate from ADIC was revealed by coinjections with authentic anthranilate; the SgcG product was found to be distinct. MS, NMR and UV-vis analysis of the SgcG product afforded data consistent with SgcG-catalyzed formation of 3-*O*-enolpyruvylanthranilic acid (OPA, **12**) from ADIC (Figure 7). Subsequent studies revealed the absolute requirement of FMN as the sole cosubstrate and that the rate of OPA formation is SgcG concentration dependent. Hence SgcG was established as an ADIC dehydrogenase whose activity complements that of SgcD in producing the C-1027 benzoxazolinate precursor OPA (Figure 7).

By elucidating the function of SgcD and SgcG, the biosynthetic pathway of the benzoxazolinate moiety leading to C-1027 is now known to proceed through OPA, a metabolite that to our knowledge has not been observed in biosynthetic pathways. Of significance, the functional assignment of these two enzymes has allowed us to formulate a pathway to benzoxazolinate production; subsequent steps require amide bond formation (by SgcD5), hydroxylation (by SgcD3), *O*-methylation (by SgcD4), thio-esterification (by SgcH), and attachment to the C-1027 enediyne moiety (by SgcD6) (*22*). The overall pathway proceeds with retention of the enolpyruvoyl moiety of chorismate as a vinyl ether resulting from SgcD's apparent evolution as a lyase-deficient ASI homolog. Because this is not observed for other metabolites originating from chorismate, the biosynthesis of C-1027 involves a unique branch point in chorismate metabolism further illuminating the evolution of chorismate processing in both primary and secondary metabolic pathways.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

The Role of SgcC4 in β-Tyrosine Synthesis

One fascinating structural feature of C-1027 is the unique *β*-amino acid moiety, (*S*)-3-chloro-4,5-dihydroxy-*β*-phenylalanine (circled in Figure 1b). Early on, we envisioned this nonproteinogenic amino acid to be derived from L-tyrosine, although conversion of L-tyrosine to this halogenated β-amino acid would require at least three enzymatic activities inclusive of a 1, 2 amine migration step. Naturally occurring *β*-amino acids are relatively rare and fall into two metabolically distinct categories (*25*). *β*-Alanine, *β*-leucine, and *β*-lysine are all known in primary metabolism. *β*-Alanine occurs in the biosynthesis of pantotheine, while the latter two play a role in the catabolism of their respective proteinogenic precursors (*25*). On the other hand, *β*-arginine, *β*-tyrosine, and *β*-phenylalanine have all been isolated as building blocks in the biosynthesis of secondary metabolites (*25*). In almost all cases, the *β*-amino acid is formed from an intramolecular migration of the α-amino group. At the time of our initial investigations of SgcC4 there were four known routes by which such an amino group migration might be enzymatically accomplished.

The four classes of aminomutases that were known at the outset of SgcC4 studies included: radical-based enzymes using either cobalamin or *S*-adenosylmethionine plus an iron-sulfur cluster, pyridoxamine phosphate-dependent enzymes, and ATP-dependent enzymes constitute three of these (*26*). Finally, the fourth mechanistic class was represented by tyrosine 2,3-aminomutase (TAM), isolated from the Edeine A producer *Bacillus brevis* Vm4 (*25*). The enzyme shows no dependence on *S*-adenosylmethionine, adenosyl-cobalamin, or pyridoxal phosphate; however, there is an absolute requirement for ATP, implying a fundamentally different mechanism from those of the radical-dependent or pyridoxal phosphate-dependent enzymes.

Analysis of the C-1027 biosynthetic gene cluster revealed that none of the genes identified encoded proteins bearing significant homology to any of the 4 classes of known aminomutases. Instead, *sgcC4* was found to encode a protein homologous to several members of a well-studied family of ammonia lyases and possessing the trademark alanine-serine-glycine (ASG) motif, characteristic of many ammonia lyases (*27*). SgcC4 homologs were found to include the histidine ammonia lyase from *Streptomyces griseus* (39% identity and 56% similarity) and the phenylalanine ammonia lyase from *Streptomyces maritimus* (38% identity and 56% similarity) (*25*). The highly conserved ASG motif in ammonia lyases is required for formation of the 4-methylideneimidazole-5-one (MIO) prosthetic group. These lyases catalyze the elimination of ammonia from α-amino acids to yield α, *β*-unsaturated acids and ammonia as products (*18*). Divided into two subfamilies on the basis of substrate specificity for either histidine or phenylalanine (*27*) it was, at the time, unprecedented for an MIO-dependent enzyme to act as an aminomutase. However, *β*-elimination of ammonia from an amino acid substrate, the reaction catalyzed by MIO-dependent ammonia lyases, accomplishes the more difficult half of the overall aminomutase transformation. Instead of releasing ammonia and the α,*β*-unsaturated acid as free products, Michael addition between the two products of the lyase reaction was envisioned to afford the *β*-amino acid as an end product

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

(*25,26*). Such a Michael reaction would clearly be favored by an enzyme similar to known ammonia lyases but lacking the same degree of active site solvent accessibility.

Extensive characterization of SgcC4 activity, indeed, confirmed that this enzyme converts α-tyrosine to the corresponding β-tyrosine **14**, a crucial precursor to the (*S*)-3-chloro-4,5-dihydroxy-*β*-phenylalanine moiety found in C-1027 (Figure 8) (*26*). Prior to our studies with SgcC4 only one example of a tyrosine-specific ammonia lyase had been reported; this from *Rhodobacter capsulatus*, where *p*-hydroxycinnamic acid (HCA), the chromophore of a photoactive yellow protein, is the product of enzymatic action (*28*). Not surprisingly, shortly after our initial identification of SgcC4, we determined that HCA, also is an intermediate in the SgcC4-catalyzed production of β-tyrosine from L-tyrosine. Rigorous kinetic analysis of SgcC4 and a S153A mutant confirmed the role of the purported MIO prosthetic group and was used to dissect the reaction path from L-tyrosine to β-tyrosine (*25*). Also determined was that SgcC4 exhibits a β-tyrosine racemase activity but is incapable of L-tyrosine racemization. These findings, well prior to the solution of the SgcC4 crystal structure, helped to establish a strong link between SgcC4 and known aminomutases. In contrast to the case of SgcD, which appears to be an ASI homolog whose lyase activity has been evolutionarily ablated, SgcC4 appears to represent a tyrosine ammonia lyase for which a mutase function, the result of Michael addition of ammonia to the product of lyase activity, has evolved "in".

SgcC4 represents a novel class of aminomutase not dependent on metals or complex cofactor chemistry and, as such, sheds light on the possible evolution of this general class of enzymes. The strong sequence and structural homology between MIO-based aminomutases and ammonia lyases suggests a common catalytic pathway. Predicated on this it has been reasoned that SgcC4, or any MIO-based L-tyrosine aminomutase for that matter, must perform the lyase chemistry followed by re-addition of ammonia stereospecifically into the coumarate intermediate to give (*S*)-β-tyrosine (**14**, Figure 8). The second half of the proposed reaction scheme is mechanistically challenging since the HCA intermediate **15** (Figure 8) is a poor electrophile and the enzyme must also retain ammonia in the neutral free base state to attain sufficient nucleophilicity. Given the high sequence homology between SgcC4 and MIO-dependent ammonia lyases, one wonders what the evolutionary relationship between SgcC4 and ammonia lyases is and whether the aminomutase activity could be engineered into ammonia lyases (and vice versa). One also wonders how widely MIO-dependent aminomutases are distributed in other organisms and what the metabolic advantages of having yet another type of aminomutase might be. Significant insight into the structural differences and similarities between SgcC4 and the tyrosine ammonia lyase (*Rs*TAL) has provided a glimpse of how SgcC4 may have evolved from the ammonia lyases. Such progress has been hastened by the recent solution of the X-ray crystal structure of SgcC4 and careful comparison to the L-tyrosine ammonia lyase *Rs*TAL from *Rodobacter sphaeroides* (*29*).

SgcC4 and the tyrosine ammonia lyase *Rs*TAL are, not surprisingly, very homologous in both primary sequence and structure (*29*). The substrate for the two enzymes is the same and the majority of side chains proximal to the active

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

site are approximately in the same location with overlap being independent of the presence or absence of substrate or analogs. Only the residue at position 415 in SgcC4 differs significantly in the region of the active site proximal to the MIO relative to the *Rs*TAL case; a very high degree of structural similarity is otherwise noted for the two enzymes. Taken together, comparative analysis of the SgcC4 and *Rs*TAL active sites suggests a highly conserved binding mode for L-tyrosine.

Differences between SgcC4 and *Rs*TAL, and perhaps many tyrosine lyases, are evident in loops surrounding the active site (*29*). In both enzymes, these loops may serve the role of shielding the reactive MIO functionality, a result in SgcC4 of self-condensation of Ala152, Ser153, and Gly154. Within SgcC4, two loops combine to shield the active site from bulk solvent in a fashion very similar to the "closed" active sites common in ammonia lyase structures. The loops in SgcC4 are more prominent than in the lyase structures with each loop contributing one residue capable of blocking off the active site (*29*). Overall, the function of the active site loops in SgcC4 may be to prevent ammonia from leaving the active site, a minimal prerequisite to differentiate an aminomutase from an ammonia lyase and suggesting that perhaps, SgcC4 may have evolutionarily preceeded the tyrosine ammonia lyases. Indeed comparison of the active sites of SgcC4 with *Rs*TAL (and other ammonia lyases), reveals that the SgcC4 active site lacks accessibility to the solvent as present in ammonia lyases. The active sites of ammonia lyases are accessible to bulk solvent through an opening in the loops forming a small channel (Figure 9b). Solvent accessibility, and hence facile diffusion of ammonia and purported Michael acceptor **15** (Figure 8) away from the MIO-bearing active site is not apparent in SgcC4 (Figure 9). The channel clearly apparent in cross sectional views of *Rs*TAL, is absent in SgcC4 by virtue of intraprotein association between two key residues (*29*). In SgcC4, but not the ammonia lyases, Tyr303 and Glu71 form a hydrogen-bonding interaction across the channel. In ammonia lyases the MIO cofactor is visible from the protein surface through a small channel. This likely hastens diffusion of lyase products on a timescale incompatible with subsequent active site chemistry. Similar views of the SgcC4 structure suggest an enclosed active site, which favors "post lyase" chemistry. In addition, the inner and outer loops of SgcC4 contain additional secondary structure elements not present in ammonia lyase structures, further shielding the active site with ordered loops. It is clear that the SgcC4 active site is significantly more closed than that of the ammonia lyases and that this structural feature likely imparts upon SgcC4 its aminomutase function. But is this really the sole explanation for SgcC4's activity deviation from bioinformatic prediction?

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

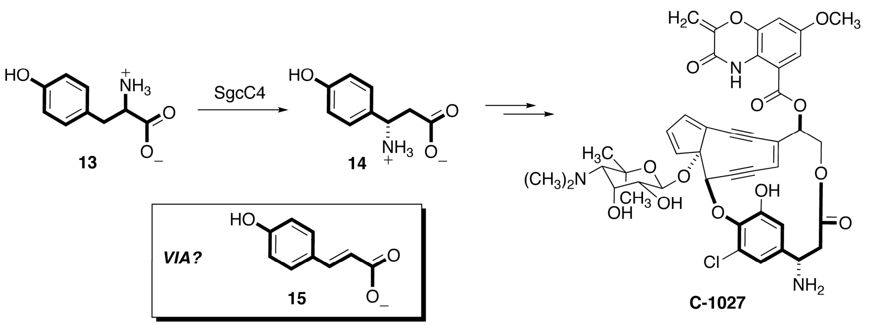


*Figure 8. Initially envisioned biosynthetic pathway β-amino acid component of C-1027 invoking SgcC4 as an aminomutase capable of converting L-tyrosine into β-tyrosine. Central to the hypothesis was transient production of p-hydroxycinnamic acid (HCA, **15**).*

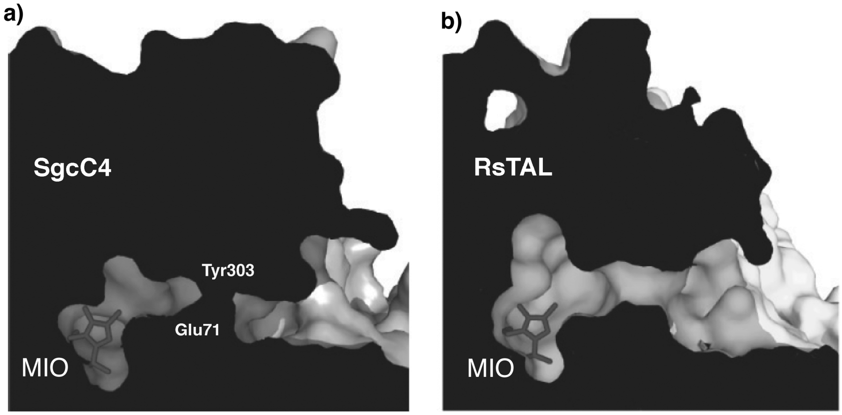


Figure 9. Cross-sectional views highlighting active site access/solvent accessibility of SgcC4 and RsTAL. The active site of SgcC4 is significantly more encapsulated than that of RsTAL by virtue of Tyr303 and Glu71.

Not yet considered in our discussion of SgcC4 is the role played by MIO. This prosthetic group is a protein-derived electrophilic moiety responsible for acidification of the benzylic protons in L-tyrosine. Two mechanisms have been proposed to account for the MIO-catalyzed elimination of ammonia observed in the ammonia lyases; both mechanisms are potentially applicable also to SgcC4 (*28,29*). One proposal invokes a Friedel-Crafts type alkylation of the aromatic tyrosine sidechain to afford the cation **16**, which undergoes conversion to HCA (**15**) through the agency of **17** (Figure 10a). Alternatively, HCA formation can be envisioned as in Figure 10b wherein the substrate undergoes N-alkylation to afford a quaternarized intermediate **18**, which is highly susceptible to elimination affording HCA and the enzyme-bound primary amine **19**. Both mechanistic options provide a viable route to ultimate formation of β-tyrosine. Significant however, is that of these two mechanistic options, amino acid N-alkylation by MIO is the path most likely to allow for subsequent Michael addition upon the lyase product HCA; the amine moiety to be incorporated within the β-tyrosine product of SgcC4 is localized within the active site and never afforded the opportunity for diffusion to solvent beyond the enzyme surface. Recent experiments using a novel difluorinated analog of L-tyrosine have permitted partial resolution of this mechanistic debate, at least for SgcC4, and make clear that reasons other than simple active site enclosement account for the enzyme's unique aminomutase activity (*30*).

The recently solved co-crystal structures of α-difluoro-β-tyrosine and its 4-methyl ether analog bound by SgcC4 provide significant insight into the chemistry catalyzed by SgcC4 and supports earlier modeling studies (Figure 11a) with L-tyrosine (*29,30*). This view of a unique inhibitor-enzyme complex unveils important details about SgcC4 activity. First and foremost, it is apparent that the β-amino moiety of both difluoro-amino acids is subject to N-alkylation by the MIO prosthetic group (Figure 11b,c). This, combined with the previously established reversibility of the SgcC4-catalyzed 1, 2 amino shift therefore strongly suggests the viability of an amino-MIO mechanism *en route* to β-tyrosine formation as depicted in Figure 10b. The apparent ability of the SgcC4 MIO to serve as an amino group "chaperone" by undergoing nucleophilic attack by the α-amine of L-tyrosine and then serving as an ammonia surrogate **19** (Figure 10b) explains, in large part, SgcC4's aminomutase activity.

In addition to covalent attachment between the β-amino acid and MIO moiety, key substrate specificity determinants of SgcC4 are apparent (*29,30*). His93 and Tyr415 form a bifurcated hydrogen bond with the substrate phenol (Figure 11b,c). This is significant since residues corresponding to His93 have previously been implicated as a key substrate determinant in tyrosine and phenylalanine ammonia lyases. Examination of the α-difluoro-β-tyrosine-SgcC4 co-crystal also reveals that the substrate carboxylate forms hydrogen bonds with Arg311 and Asn205 (Figure 11b,c). In general, the active site is characterized by a number of aromatic amino acids, perhaps the most important of which is Tyr63. Tyr63 is positioned in close proximity to the α and β-protons of the inhibitor and is well-positioned to assist in elimination chemistry (*30*). Gly70 forms a hydrogen bond with the phenol of Tyr63 and both residues are conserved in all MIO-based lyases and mutases. With the assistance of Gly70, the phenolate of Tyr63 is proposed to serve as a general base facilitating elimination (after amino group activation as a leaving group) and the ensuing formation of HCA (**15**, Figure 10). H-bonding of the Tyr63 phenol with Gly70 favors phenolate formation as does the abundance of R-helices with positive dipoles pointed at the active site. Also supportive of this role for Tyr63 is the observed pH optimum for SgcC4 activity of ~9.2, and that replacement of Tyr63 with Phe ablates aminomutase activity. These data support the modeling studies that depict the phenolate of Tyr63 as being able to deprotonate either the α or β protons of L-tyrosine (*29,30*). Moreover, these studies explain, in part, the fact that SgcC4 displays β-tyrosine racemase activity. In sum, extensive structural studies with SgcC4 and a novel difluorinated β-tyrosine derivative support the importance of select amino acids within the active site and the similarity of these residues to those found in other ammonia lyase enzymes. Most important though is that these studies firmly implicate an amine-MIO intermediate (**19**, Figure 10b) as being crucial to SgcC4's activity as an aminomutase and not a simple lyase.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

That SgcC4 is an aminomutase is counter to bioinformatic predictions. The enzyme is most similar to known amino acid lyases and thus initial findings revealing the enzyme's inability to produce the HCA as a final product from L-tyrosine, were remarkable. Despite apparent similarities of structure and function between SgcC4 and known ammonia lyases two characteristics of SgcC4 appear to equip this ammonia lyase with the additional ability to catalyze formation of an N-βC bond. Limited solvent accessibility of the active site along with the use of an amino-linked-MIO intermediate (Figure 10b) are both unique features of SgcC4 that enable chemistry beyond the lyase reaction. It is difficult to say at this juncture if SgcC4 is an evolutionary remnant of the ammonia lyases or if this enzyme is at the forefront of how the amino acid lyases may be continually evolving. Regardless of the enzyme's evolutionary relationship to closely related enzymes, what is clear, is that our understanding of SgcC4 and its relationship to other aminomutases and lyases creates tremendous opportunity for rational engineering of new natural products via targeted modifications to these amino acid modifying enzymes.

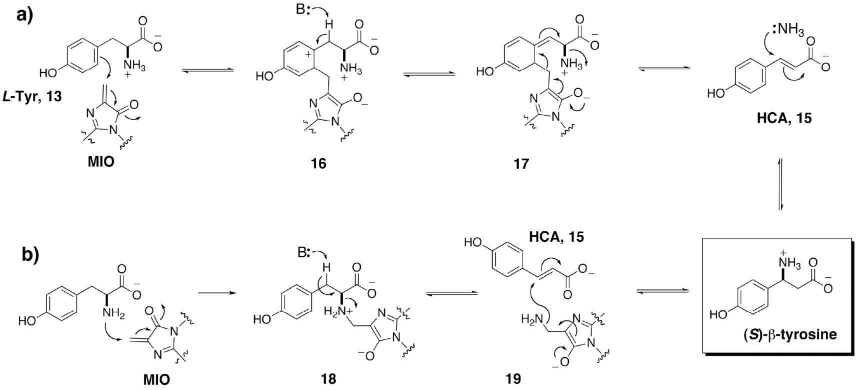


Figure 10. Mechanistic proposals for MIO-dependent aminomutase chemistry based on the proposed mechanistic schemes for MIO-based ammonia lyases. (a) Friedel-Crafts mechanism invoking acidification of benzylic protons following carbocation formation. (b) Amino-MIO adduct mechanism invoking α-amino group quaternarization via alkylation with MIO followed by C-N bond scission and subsequent MIO-adduct directed Michael addition to afford the new C-N bond.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

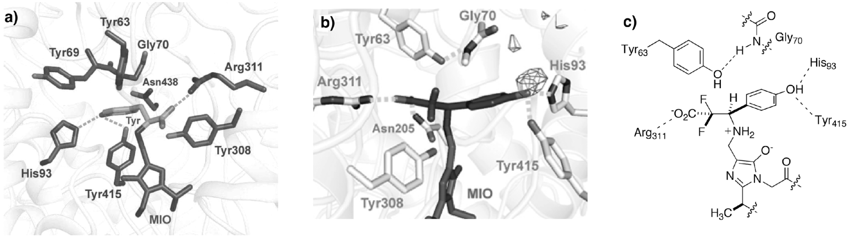


Figure 11. Views of the SgcC4 active site associated with substrate and substrate analogs. (a) Model of the substrate L*-tyrosine bound in the active site of SgcC4. (Adapted with permission from reference 29, copyright 2007 American Chemical Society) (b) View of the SgcC4 active site showing the electron density map calculated from co-crystals of α-difluoro-β-4-methoxyphenylalanine showing the position of the methyl group and allowing correlation to the SgcC4 structure bound to α-difluoro-β-tyrosine (30). (c) Diagramatic representation of the major interactions and conformation of the enzyme with the bound α-difluoro-β-tyrosine. (Adapted with permission from reference 30. Copyright 2007 American Chemical Society.)*

Production of New Bioactive C-1027 Analogs via Combinatorial Biosynthesis: Evolutionary Applications of Natural Products Biosynthesis

"Combinatorial biosynthesis" – in its broadest sense, the generation of novel analogs of natural products by genetic engineering of biosynthetic pathways – complements total synthesis and semi-synthesis which often bear the burden of high production and environmental costs, particularly as the molecular complexity of target molecules increases (*5,31-33*). The prospect of producing new molecules by modifying existing pathways – or even building new pathways from scratch – via such combinatorial biosynthesis methods represents a revolutionary concept and constitutes an important evolutionary application of recombinant DNA technology, particularly in *Streptomyces* and related microorganisms known for their ability to produce natural products.

The mechanism of antitumor action of C-1027, as with all enediynes, is most closely associated with O_2-dependent DNA strand scission resulting from cycloaromatization of the enediyne component. Cycloaromatization produces carbon-centered radicals that abstract H-atoms from the sugar phosphate backbone. In the C-1027 case, H-atom abstraction proceeds with greatest efficiency at the deoxyribose 1', 4' and 5' centers; the carbon-centered radical in each case rapidly reacts with O_2 to afford highly labile oxidized sugars that lead to strand breaks (*34*). Cells typically respond to enediyne-induced damage as they would to ionizing radiation induced damage. An interesting feature of C-1027 not displayed by most other enediyne natural products is its ability to induce not only DSBs but also interstrand crosslinks (ICLs), particularly under O_2-depleted conditions where a conventional pathway to sugar oxygenation becomes restricted (*35*). C-1027 induces ICL formation not only under cell-free anaerobic conditions but also in cells (*35*). That C-1027 induces both DSBs and ICLs under aerobic conditions suggests that C-1027 might provide a unique opportunity for the application of combinatorial biosynthetic methods. Specifically, it was envisioned that C-1027's mechanism of action might be fine-tuned so as to afford an agent with reduced activity in the presence of O_2 yet retaining DNA damaging activity under hypoxic conditions. Such a departure from the C-1027 activity profile might prove highly useful since it is well known that many classes of solid tumors establish hypoxic regions that provide unique opportunities for therapeutic targeting selectivity.

Sequencing of the C-1027 biosynthetic gene cluster and many subsequent efforts indicated that oxidation and chlorination at both positions ortho to the tyrosine-derived phenol (C20' and C22', Figure 12a) are catalyzed by the gene products coded for by *sgcC3* and *sgcC*, respectively (*15*). Additionally, it was determined that *sgcD4* codes for a methyltransferase responsible for methylation of the benzoxazolinate phenol moiety (*15*). Using this knowledge of the C-1027 biosynthetic machinery, combinatorial biosynthesis methods were applied to abolish the function of *sgcC*, *sgcC3* and *sgcD4* individually via targeted gene inactivation (Figure 12a) (*35*). Inactivation of *sgcD4* provided a mutant strain of *S. globisporus* that produced 7''-desmethyl-C-1027 (**20**), whereas similar inactivations of *sgcC* and *sgcC3* afforded 22'-deshydroxy-C-1027 (**21**), and 20'-deschloro-C-1027 (**22**), respectively (Figure 12a).

Initial assessment of bioactivities displayed by the C-1027 analogs **20**, **21** and **22**, obtained by combinatorial biosynthesis (relative to C-1027) emphasized the impact of these agents upon cell growth of YZ510B human fibroblast cells (*36*). Continuous subjection to C-1027 and related analogues for 3 days resulted in measured IC_{50}s for growth arrest of 24 pM for C-1027, 1,412 pM for **20**, 174 pM for **21**, and 72 pM for **22**, which is approximately 60-, 7-, and 3-fold higher than C-1027, respectively (Figure 12b) (*36*). A profound reduction in growth inhibition was apparent for **20**, the C-1027 analog missing the benzoxazolinate methyl ether moiety (circled, Figure 1b).

The significant differences in cell growth inhibition among C-1027 analogs produced with knowledge of the biosynthetic machinery warranted significantly greater investigation. As highlighted in Figure 12c all C-1027 analogs induced DSBs although normalization to C-1027 revealed that the desmethyl analog **20**, the deshydroxy analog **21** and deschloro analog **22**, are 50, 30, and 4 times less effective than C-1027 (*35*). To determine the extent to which these C-1027 analogs induced ICLs under anaerobic cell-free conditions, compound-treated, linearized plasmid DNA was used to evaluate the amount of double-stranded DNA remaining after alkaline denaturation; crosslinked DNA displays electrophoretic mobility discernibly different from that of the non-crosslinked, readily denaturable duplex. At 13 nM C-1027, ~13% dsDNA remained, which increased to ~30% by 130nM. As shown in Figure 12c, desmethyl analog **20** induced ICLs with an efficiency on par with C-1027 whereas the deschloro and deshydroxy-C-1027 analogs displayed severely attenuated interstrand crosslinking capabilities (*35*). Indeed, ICLs resulting from treatment with deshydroxy analog **21** were not readily detected. It is perhaps highly significant that follow up studies using intracellular SV40 DNA as a substrate substantiate

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

the data resulting from cell free studies (Figure 12c). At 8 nM, C-1027 induced significant amounts of linear intracellular SV40 DNA (20%), which increased in a concentration-dependent manner to ~40%. The deschloro analog also induced DSBs in a concentration-dependent manner although ~8-fold higher drug concentrations (relative to C-1027) were required. The desmethyl analog initially induced a minimal amount of intracellular SV40 DSBs, but additional breaks were not observed even at concentrations as high as 10 μM, which is also consistent with previous genomic DNA studies (*36*). The potency of the C-1027 family members at inducing intracellular SV40 DSBs is C-1027 > deschloro >> desmethyl. Finally, both C-1027 and its desmethyl analog **20** were found to induce ICLs in intracellular SV40 while no such adducts were found upon treatment of cells with the deschloro analog **22**. The data from cell free and intracellular SV40 DNA-based assays have been further substantiated by studies to evaluate the impact of C-1027 and its desmethyl and deschloro analogs upon genomic DNA in HCT116 tumor cell lines.

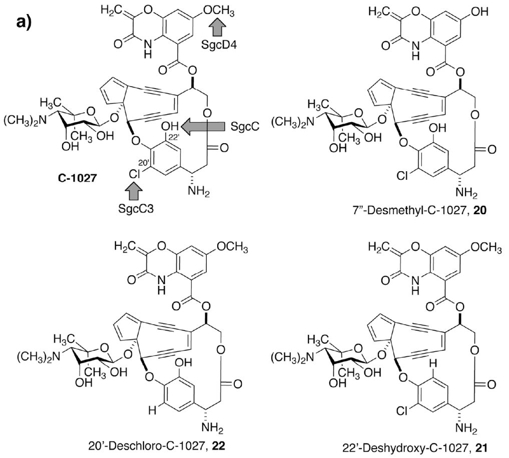
For C-1027, concurrent induction of DSBs and ICLs provides a unique opportunity to test the paradigm that cells require the protein kinase ataxia-telangiectasia mutated (ATM) to activate DNA damage responses to strand breaks, whereas the kinase ATM and Rad3-related (ATR) is required to respond to ICLs (*35,36*). The ability of combinatorial biosynthesis to readily afford C-1027 analogs capable of DSB induction but not ISC induction (deschloro and to a much lesser extent deshydroxy **21**) and of analogs capable of almost exclusive ICL induction (desmethyl analog **20**) has therefore been used to great advantage to better understand the role of ATM and ATR in cellular responses to radiomimetic drugs such as C-1027.

At present, determining whether a radiomimetic drug such as C-1027 has the potential to induce ICLs is not readily apparent. Although all enediynes induce DSBs under aerobic conditions, ICL production under anaerobic conditions requires that deoxyribose radicals generated upon H-atom abstraction back react with the drug (*34*). The efficiency of ICL induction, which is thought to depend on the proximity and steric availability of the enediyne chromophore to the deoxyribose radicals, differs greatly among enediynes, yet the variation in DNA break to cross-link activity between C-1027 and its analogs was unexpected given their minimal structural differences (Figure 12a). Detailed hypotheses have been put forth that take into account both the impact of steric and reactivity differences between C-1027, **20**, **21**, and **22** (*35*). Although the precise mechanistic reasons for different bioactivities of C-1027, **20**, **21** and **22** remain unclear, these findings highlight the potential for rational biosynthetic engineering to yield new generations of antitumor compounds able to produce cellular DSBs, ICLs, or a combination of both. This is but one example of how the field of biosynthesis is evolving to exploit the advances in molecular biology to produce and discover new natural products with tremendous potential. In this specific context it is remarkable that such small structural differences among C-1027 analogs translate to such significant differences in activity. These structural differences shift cellular responses to DNA damage from ATM-dependent to ATR-dependent or can permit use of both kinases in accordance with the proportion of DSBs to ICLs (*36*). Remarkably, these minor modifications to the enediyne chromophore (relative to C-1027) dramatically

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

alter DNA damage patterns, suggesting a rational approach for the design of new generations of antitumor agents.

This work shows that specific bioactivities displayed by the parent compound can be evolved out to afford agents with more highly refined DNA damaging activity using combinatorial biosynthesis. C-1027, with its dual mode of DNA damage induction (Figure 13) is regarded as a radiomimetic agent in that it induces O_2-dependent DSBs. However, it also induces O_2-independent ICLs. That the desmethyl C-1027 analog **20** retains C-1027's capacity for ICL formation but lacks the parent compound's capacity for O_2-dependent DSB induction may have a profound impact on this analog's consideration as a chemotherapeutic. That many solid tumors establish O_2-depleted or hypoxic regions (Figure 13) might represent a means by which **20** can express more highly selective and less host-toxic DNA damaging activity than is displayed by C-1027 (*35*). It is clear that combinatorial biosynthesis constitutes what is essentially "intelligent evolution" and that the basic science behind it represents a critical step in the evolution of natural products biosynthesis. However, the results of biological testing of C-1027 and its engineered analogs **20**, **21**, and **22** highlight how the evolution of studies in biosynthesis may impact drug discovery.



b)

compound	*IC$_{50}$ (pM)*	*Activity relative to C-1027*
C-1027	24	1
7"-Desmethyl-C-1027 (**20**)	1.4×10^3	60
22'-Deshydroxy-C-1027 (**21**)	1.7×10^2	7
20'-Deschloro-C-1027 (**22**)	72	3

c)

compound	*DSB induction relative to C-1027*	*ICL induction relative to C-1027*
C-1027	1	1
7"-Desmethyl-C-1027 (**20**)	0.02	1
22'-Deshydroxy-C-1027 (**21**)	0.03	ND
20'-Deschloro-C-1027 (**22**)	0.25	< 0.003

*Figure 12. C-1027, engineered analogs and associated biological activities/efficiencies. (a) Correlation of biosynthetic gene products and C-1027 functional groups. Shaded arrows indicate functional groups resulting from highlighted gene products. Structures resulting from ablation of SgcD4, SgcC and SgcC3 are indicated by **20**, **21**, and **22**, respectively. (b) Summation of C-1027 and related analog activity against YZ510B cells. (Reproduced with permission from reference 36. Copyright 2007 Cancer Research.) (c) Comparison of C-1027 and analog-induced DSBs and ICLs (35). ND, not detected.*

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

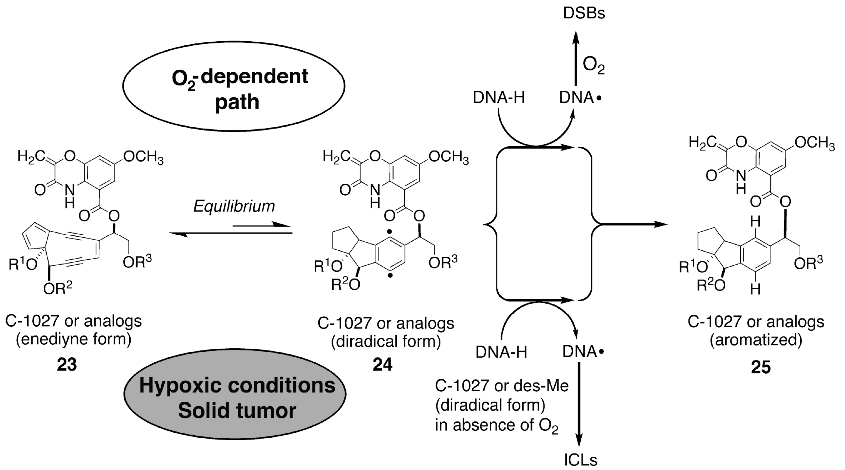

DSBs
O2
DNA-H
DNA•
O2-dependent path
Equilibrium
C-1027 or analogs (enediyne form)
23
C-1027 or analogs (diradical form)
24
C-1027 or analogs (aromatized)
25
Hypoxic conditions Solid tumor
DNA-H
DNA•
C-1027 or des-Me (diradical form) in absence of O2
ICLs

Figure 13. Overview of different DNA lesions inducible by C-1027 and related analogs based on extent of oxygenation, which differs between normoxic (non-cancerous) and hypoxic conditions typical in many solid tumors.

Closing Thoughts

It is clear from our studies of C-1027 biosynthesis and the biosynthesis of many other enediyne-containing natural products that snapshots of enzyme evolution are extraordinarily abundant. Studies of SgcD and SgcC4 are particularly striking examples of this where bioinformatic predictions and a clear relationship to other characterized and well understood enzymes are nullified by the apparent losses and gains of function of each enzyme, respectively. A superficial description of SgcD would detail it as an anthranilate synthase that has lost its chorismate lyase activity. Similar descriptions of SgcC4 would highlight this biosynthetic enzyme as an ammonia lyase that has gained the ability to also catalyze formation of a new N-C bond resulting in a β-amino acid. Both enzymes, and their characterization, clearly shed light on the evolution of whole classes of enzymes although at present it is difficult to ascribe these enzymes as evolutionary remnants or harbingers of new enzyme classes not yet discovered. Conversely, the PKSEs sequenced to date adhere closely to bioinformatic predictions of function and as such, are highly unique enzymes that serve as excellent beacons for the discovery of new enediynes. This, coupled with advances in genome mining, promises to unveil untold numbers of new natural products from previously unrecognized enediyne producers highlighting how the evolution of the PKSE has directly impacted drug discovery strategies. Finally, our understanding of biosynthetic enzymes with a high level of resolution enables combinatorial biosynthesis as a means by which to produce new "non-natural" natural products displaying novel, and potentially exploitable activity differences from those of the more well studied parent natural products. Metabolic engineering of biosynthetic pathways represents an evolutionary application extremely well suited to drug discovery efforts. Though

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

not an exhaustive list, these biosynthetic endeavors support the idea that enediyne biosynthesis is an outstanding arena for study through an evolutionary lens.

Acknowledgement

Studies on enediyne biosynthesis and anticancer drug discovery described from the Shen laboratory were supported in part by NIH grants CA78747, CA113297, and A151689. We are indebted to a group of dedicated co-workers for their many efforts as indicated in the cited references, to our expert collaborators Prof. Steven D. Bruner (Boston College), Prof. Michael D. Toney (University of California-Davis), Dr. Terry Beerman (Roswell Park Cancer Institute), and to Dr. Yi Li (Institute of Medical Biotechnology, Chinese Academy of Medical Science) for the C-1027 producing *Streptomyces globisporus* wild-type strain.

References

1. Fischback, M. A.; Walsh, C. T.; Clardy, J. *Proc. Natl. Acad. Sci. USA* **2008,** *105,* 4601–4608.
2. Lawrence, J. G.; Roth, J. R. *Genetics* **1996,** *143,* 1843–1860.
3. Firn, R. D.; Jones, C. G. *Mol. Microbiol.* **2000,** *37,* 989–994.
4. Firn, R. D.; Jones, C. G. *Nat. Prod. Rep*. **2003,** *20,* 382–391.
5. Van Lanen, S. G.; Shen, B. *Curr. Top. Med. Chem*. **2008,** *8,* 448–459.
6. Liu, W.; Ahlert, J.; Gao, Q.; Wendt–Pienkowski, E.; Shen, B.; Thorson, J. S. *Proc. Natl. Acad. Sci. USA* **2003,** *100,* 11959–11963.
7. Davies, J.; Wang, H.; Taylor, T.; Warabi, K.; Huang, X. –H.; Anderson, R. J. *Org. Lett.* **2005,** *7,* 5233–5236.
8. Nicolaou, K. C.; Smith, A. L.; Yue, E. W. *Proc. Natl. Acad. Sci. USA* **1993,** *90,* 5881–5888.
9. Gu, F.; Xi, Z.; Goldberg, I. H. *Biochemistry* **2000,** *39,* 4881–4891.
10. Sugiura, Y.; Arakawa, T.; Uesugi, M.; Shiraki, T.; Ohkuma, H.; Konishi, M. *Biochemistry* **1991,** *30,* 2989–2992.
11. Nicolaou, K. C.; Dai, W. –M. *Angew. Chem.* **1991,** *103,* 1453–1481.
12. Gleiter, R.; Kratz, D. *Angew. Chem.* **1993,** *105,* 842–845.
13. McGlinchey, R. P.; Nett, M.; Moore, B. S. *J. Am. Chem. Soc.* **2008,** *130,* 2406–2407.
14. Oh, D. –C.; Williams, P. G.; Kauffman, C. A.; Jensen, P. R.; Fenical, W. *Org. Lett*. **2006,** *8,* 1021–1024.
15. Liu, W.; Christenson, S. D.; Standage, S.; Shen, B. *Science* **2002,** *297,* 1170–1173.
16. Zhang, J.; Van Lanen, S. G.; Ju, J.; Liu, W.; Dorrestein, P.; Li, W.; Kelleher, N. L.; Shen, B. *Proc. Natl. Acad. Sci. USA* **2008,** *105,* 1460–1465.
17. Jiang, H.; Zirkle, R.; Metz, J. G.; Braun, L.; Richter, L.; Van Lanen, S. G.; Shen, B. *J. Am. Chem. Soc.* **2008,** *130,* 6336–6337.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

18. Zazopoulos, E.; Huang, K.; Staffa, A.; Liu, W.; Bachmann, B. O.; Nonaka, K.; Ahlert, J.; Thorson, J. S.; Shen, B.; Farnet, C. M. *Nat. Biotech.* **2003,** *21,* 187–190.
19. Ahlert, J.; Shepard, E.; Lomovskaya, N.; Zazopoulos, E.; Staffa, A.; Bachmann, B. O.; Huang, K.; Fonstein, L.; Czisny, A.; Whitwam, R. E.; Farnet, C.; Thorson, J. S. *Science* **2002,** *297,* 1173–1176.
20. Van Lanen, S. G.; Oh, T. –J.; Liu, W.; Wendt–Pienkowski, E.; Shen, B. *J. Am. Chem. Soc.* **2007,** *129,* 13082–13094.
21. Shen, B. *Curr. Op. Chem. Biol.* **2003,** *7,* 285–295.
22. Van Lanen, S. G.; Lin, S.; Shen, B. *Proc. Natl. Acad. Sci. USA* **2008,** *105,* 494–499.
23. He, Z.; Stigers Lavoie, K. D.; Bartlett P. A.; Toney M. D, *J. Am. Chem. Soc.* **2004,** *126,* 2378–2385.
24. He, Z.; Toney, M. D. *Biochemistry* **2006,** *45,* 5019–5028.
25. Christenson, S. D.; Weiming, W.; Spies, A.; Shen, B.; Toney, M. D. *Biochemistry* **2003,** *42,* 12708–12718.
26. Christenson, S. D.; Liu, W.; Toney, M. D.; Shen, B. *J. Am. Chem. Soc.* **2003,** *125,* 6062–6063.
27. Excellently reviewed in Poppe, L.; Retey, J. *Angew. Chem. Int. Ed.* **2005,** *44,* 3668–3688.
28. Kyndt, J. A.; Meyer, T. E.; Cusanovich, M. A.; Ban Beeumen, J. J. *FEBS Lett.* **2002,** *512,* 240–244.
29. Christianson, C. V.; Montavon, T. J.; Van Lanen, S. G.; Shen, B.; Bruner, S. D. *Biochemistry* **2007,** *46,* 7205–7214.
30. Christianson, C. V.; Montavon, T. J.; Festin, G. M.; Cooke, H. A.; Shen, B.; Bruner, S. D. *J. Am. Chem. Soc.* **2007,** *129,* 15744–15745.
31 Galm, U.; Shen, B. *Exp. Opin. Drug Discov.* **2006,** *1,* 409–437.
32. Shen, B. *Sci STKE* **2004,** *225,* 14.
33. Shen, B.; Liu, W.; Nonaka, K. *Curr. Med. Chem.* **2003,** *10,* 2317–2325.
34. Xu, Y. –J.; Xi, Z.; Zhen, Y. –S.; Goldberg, I. H. *Biochemistry* **1997,** *36,* 14975–14984.
35. Kennedy, D. R.; Ju, J.; Shen, B.; Beerman, T. A. *Proc. Natl. Acad. Sci. USA* **2007,** *104*, 17632–17637.
36. Kennedy, D. R.; Gawron, L. S.; Ju, J.; Liu, W.; Shen, B.; Beerman, T. A. *Cancer Res.* **2007,** 67*,* 773–781.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch011

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

Chapter 12

Molecular Machines

Natural and Artificial Molecular Motors

Thomas W. Bell* and Joseph I. Cline

Department of Chemistry, University of Nevada, Reno, NV 89557-0216

Artificial molecular motors are a natural extension of biological motors, which have evolved to perform many different functions in living organisms. *Homo sapiens* is further adapting biomotors to drive nanoscale devices. Artificial motors are of keen interest because they can potentially work faster than biomotors, they can be more robust, and they can be driven and controlled by light. Chemists have designed many types of chemically-powered and light-powered rotary molecular motors and some have been shown to undergo unidirectional rotation. Molecular dynamics simulations and photoisomerization studies have also underscored the feasibility of the authors' own photon-driven ratchet design.

Biological systems have evolved many, complex molecular and supramolecular devices, which enable organisms to grow, reproduce, sense stimuli, and interact with their environment. *Homo sapiens* has increasingly engaged in mimicry of natural, functional molecules, including total synthesis of biologically active natural products and creation of artificial systems capable of performing natural functions. In the authors' opinion, this produces an evolutionary continuum between natural molecules and artificial compounds that have been inspired by nature. Evolution of anthropogenic molecules is a natural extension of biological evolution. The human brain, itself a product of evolution, is inspired by marvelous natural devices to create artificial analogs that may be capable of operating even under abiologic conditions. These

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

devices are also subject to selective pressure, since designed molecules that do not function are abandoned.

The purpose of this chapter is to consider current knowledge and recent developments concerning a specific kind of molecular device, the molecular motor. Accordingly, some definitions are in order. A *molecular device,* whether natural or artificial, is commonly considered to be a molecule that can perform a useful task. Examples of molecular devices are chemosensors (*1,2*), switches (*2*), and motors. A *molecular machine* (*3-6*) is capable of doing useful work. Both molecular switches and motors can accomplish mechanical work, but the work executed by a switch is undone when it is reset to the original position. A *molecular motor* is a device that is capable of doing work repetitively and progressively, so that *net* mechanical work is performed over multiple cycles of actuation (*6*). The physical definition of work (force times distance) is applicable here. Molecular motors can be divided into two types: rotary and linear, as illustrated by examples given later. The hierarchy is clear, proceeding from the broad class of molecular devices, to the more specific molecular machines, of which rotary and linear molecular motors are particular types.

Natural Molecular Motors

Natural molecular motors (motor proteins) are intriguing, both as efficient products of evolution and as power sources for nanotechnology (*7*). Some representative rotary and linear motor proteins are subjects of this chapter. Not discussed here are a wide range of additional proteins in which linear motion is coupled with enzyme activity, or which perform physical work. Examples are RNA polymerase and DNA topoisomerase, respectively.

F_0F_1-ATP Synthase

Of particular interest with respect to applications in nanotechnology is the rotary supramolecular motor, F_0F_1-ATP synthase (Figure 1) (*8*). This enzyme is actually a double rotary motor. Proton flux across the membrane drives rotation of the F_0 unit and the γ "spindle" of the F_1 unit. Rotation of the γ spindle within the stationary F_1 unit, in turn, drives the synthesis of ATP from ADP and inorganic phosphate. While the rotary motion of this device qualifies it as a motor, this motion is merely a mechanism, enabling its biological function as a factory for a ubiquitous fuel needed by cells.

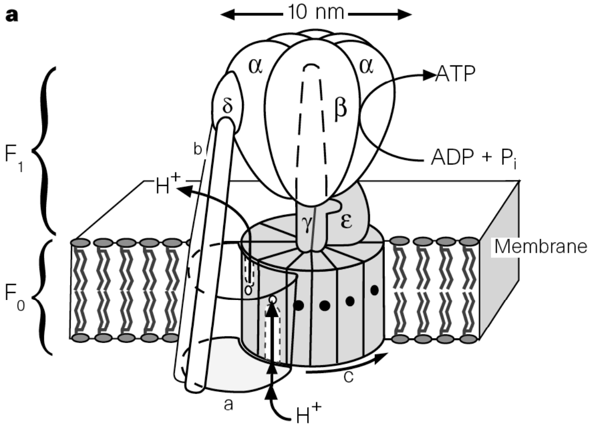


Figure 1. Structure of F_0F_1-ATP synthase. (Reproduced from reference 8. Copyright 1998 MacMillan Publishers, Ltd.).

The F_1 unit of F_0F_1-ATP synthase can be isolated and used as a chemically driven motor for nanotechnology. The F_1 motor has been immobilized on surfaces via the α/β assembly, and the protruding γ spindle has been attached to actin filaments (*9*) or to ~1 μm nickel rods (*10*). This motor catalyzes ATP hydrolysis, driving rotation of the γ spindle and its attached "propeller" in the opposite rotational sense, with respect to the normal rotary direction in ATP synthesis.

Bacterial Flagellar Motor

One of the most impressive products of evolution is the bacterial flagellar motor (Plate 1) (*11,12*). This assembly of more than 20 different proteins is imbedded in the cell wall and membrane of some bacteria, including *E. coli.* Proton flux across the membrane drives rotation of the wheel and the attached, helical flagellum, producing locomotion of the bacterium.

Interest in the mechanism of evolution of the bacterial flagellar motor stems, in part, from its current role as a poster child for the "intelligent design" community. The argument of *sub rosa* (as well as outright) creationists is that this machine exhibits "irreducible complexity." (*13*) In particular, the device could not function if even only one of its component proteins were missing. Evolution would require that all of the component proteins undergo natural selection simultaneously, absent functionality that is required for genetic optimization. Some proponents of intelligent design also argue that the amazingly precise self-assembly of so many proteins to produce an exquisitely complex machine also implicates the hand of at least a designer, if not a creator. Of course, the same thing could, and has been, said about complex organisms, such as *Homo sapiens*.

Ironically, protein self assembly can solve the problem of evolution of complex, multicomponent machines, such as the bacterial flagellar motor[14]. Chemists understand that self-assembled systems are fluctional and constantly

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

sample alternate configurations. Structural homologies are often observed between proteins having different functions. This suggests that proteins that evolved to perform one function later became capable of another function, probably on a more rudimentary level. This alternate function gave the organism a competitive advantage, leading to evolutionary optimization of the second function.

It turns out that there are extensive homologies between proteins in the basal region of the bacterial flagellum and type III secretory proteins comprising structures in the membranes of infective agents such as the plague bacterium, *Yersina pestis*. This bacterium, thought by many to be responsible for some epidemic plagues, uses type III secretory proteins to inject protein toxins into target cells of the host. Interestingly, the bacterial flagellum self assembles from proteins extruded through a central pore, and mature flagella can serve also to export proteins from the cell. Thus, a structure that evolved to export proteins may have been co-opted evolutionarily to produce locomotion (*14*).

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

Transport and Muscle Proteins

Linear molecular motors are represented in biology (*15*) by the transport proteins, kinesin and dynein (*16*), and by the muscle protein, myosin. Kinesins (Plate 2) (*17*) attach to microtubules and take several hundred “steps” before detaching. Their cargo includes various proteins, messenger RNA, and vesicles. ATP hydrolysis causes a conformational change in the arm attached to the head group that is bound to the microtubule. This enables the second head group to move “forward” and attach to the microtubule before release of the first head. Myosin (Plate 3) (*17*) also has two head groups, but only one attaches to a muscle actin filament. ATP hydrolysis drives a single forward stroke, then the head detaches, enabling neighboring motors to continue forward progression.

Crystallographic studies have revealed structural similarities between the kinesin and myosin protein families, and there are short amino acid stretches displaying sequence conservation (*15*). These findings have led to the proposal that kinesins and myosins have a common ancestor. This precursor protein apparently also led to the G protein superfamily, which has a “phosphate sensor” domain that is very similar to the ATP binding sites of kinesins and myosins[15]. As proposed for evolution of the bacterial flagellar motor, transport and locomotion functions appear to be evolutionarily connected.

Designed Molecular Motors

Potential uses of molecular motors include powering nano-assemblers, altering properties of nanomaterials, pumping fluids or disrupting laminar flow, powering nanovessels, switching nanocircuits, releasing trapped molecules (i.e. in drug delivery), and modulating functions of biomolecules, such as proteins and nucleotides. Why not employ natural molecular motors to perform these functions? After all,

- Biomotors already exist, and they work!
- They are very efficient.
- And, they use a common chemical fuel.

On the other hand:

- Biomotors lack precise control.
- They are relatively slow (10-100 Hz).
- And, they work only under limited conditions (solvent, temperature, and pH).

For highest versatility, operating under diverse conditions, artificial molecular motors should have the following properties, in the authors' opinion.

- Artificial motors should be rotary, and operate unidirectionally.
- They should be powered and controlled by light.
- They should be locked when unenergized.
- They should be capable of high rotary speed.
- They should be robust (stable under diverse conditions, and potentially operable in a vacuum).
- And, they should be adaptable, enabling the work they perform to be harnessed for useful purposes.

Many ingenious designs of molecular motors have been reported, including a light-driven family developed by Feringa and coworkers (*18*), and chemically-driven prototypes produced by Kelly (*19*). Included in this chapter is a brief discussion of recent developments in the field of artificial molecular motors, including the authors' own design.

Chemically-Powered Artificial Motors

Kelly designed a ratchet-type chemically-powered molecular motor consisting of a three-bladed triptycene gear and a helical "pawl", which both prevents back slippage and directs the motion of the gear (Plate 4) (*20*). The Kelly group first synthesized a simpler prototype lacking the 4-dimethylaminopyridyl (DMAP) group and two of the amino groups of compound **1**, shown in Plate 4. They demonstrated that reaction of the prototype with phosgene yields a strained product having a carbamate tether between the gear and the pawl. Thermal relaxation accomplishes a 120° clockwise rotation (*19*). The full motor (**1**, Plate 4) contains a DMAP group, strategically positioned to catalyze reaction of phosgene with the "correct" amino group. Formation of the carbamate tether, thermal relaxation, and hydrolysis of the tether would produce 120° rotation, returning the molecule to the form required for further reaction with phosgene. Unfortunately, this ingenious device stalled after reaction of phosgene with the "correct" amino group, producing carbamate **6** (Plate 4) (*20*).

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

Branchaud designed a structurally simpler chemically-powered molecular motor based on chemoselective formation of biaryl lactone **10** (Figure 2) (*21*), followed by chemoselective hydrolysis. The racemic lactone was synthesized, then cleaved with directional (ca. 90°) rotation of the two aromatic rings. Relactonization gave **12** (Figure 2) with very high 180° rotational selectivity, overall. Unfortunately, conditions were not discovered enabling complete, 360° rotation in this system. Feringa and coworkers previously reported a biaryl system that undergoes unidirectional rotation when subjected to a sequence of chemical reactions, including asymmetric reduction and reoxidation (*22*). Thus, it appears that chemists have not currently succeeded to demonstrate full 360° rotation in an artificial molecular motor driven by a single chemical fuel, as in the case of biomotors.

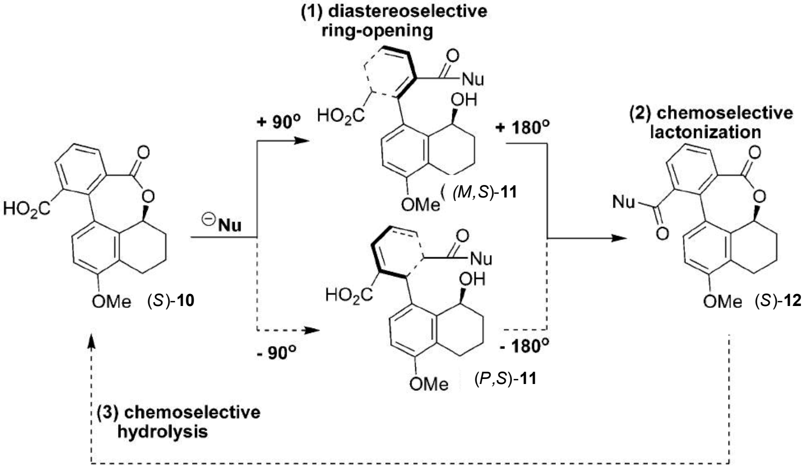

(1) diastereoselective ring-opening
+ 90°
- 90°
+ 180°
- 180°
(2) chemoselective lactonization
(3) chemoselective hydrolysis
⊖Nu
HO2C
O
OMe
Nu
OH
(S)-10
(M,S)-11
(P,S)-11
(S)-12

Figure 2. Branchaud's chemically-powered biaryl lactone rotary motor. (Reproduced from reference 21. Copyright 2006 American Chemical Society.).

Light-Powered Molecular Motors

Artificial molecular motors powered by light have been the subject of an excellent recent review (*23*). The most well developed fully functional light-powered motors have been produced by Feringa and coworkers (*18*); some second-generation systems are shown in Figure 3 (*24*). Feringa's motors are chiral, overcrowded alkenes that photoisomerize unidirectionally producing a strained product, which thermally relaxes to a lower energy form. These two steps cause 180° rotation about the central double bond. A second photoisomerization and thermal relaxation complete full, directional 360° rotation. Second-generation systems shown in Figure 3 were redesigned to improve performance, especially by lowering the barriers for the rate-determining thermal steps to increase rotary speed. While earlier models had two stereocenters, in **13** (Figure 3) unidirectional rotation is achieved with a single asymmetric carbon, and the X and Y groups can be varied to tune the

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

energy barrier. Replacing six-membered by five-membered rings (**13→14→15**, Figure 3) greatly increases the speed of rotation. Motor **16** (Figure 3), with push-pull conjugation, was engineered to operate with visible light (λ_{max} 436 nm), instead of UV. Motor **17** (Figure 3) has thiol tethers enabling its immobilization on gold nanoparticles (*24*).

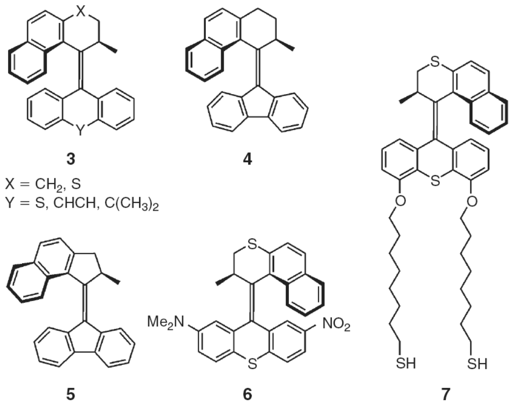


Figure 3. Feringa's second-generation light-powered molecular motors. (Partially reproduced from references 23 and 24. Copyright 2006 and 2008 A. Credi and CSIRO PUBLISHING)

Nobel Laureate Jean-Marie Lehn has conceptually reduced the light-powered motor to a very simple system: the chiral imine (Figure 4) (*25*). As observed for chiral alkenes, absorption of a photon produces out-of-plane rotation of the π bond in imines. Molecular asymmetry should cause the 90° rotation to occur with some preference for clockwise or counterclockwise direction. Photoisomerization of the anti imine to the syn geometry constitutes directional, 180° rotation. At room temperature, the syn isomer would rapidly undergo in-plane inversion to the original anti form. This brilliant simplification of the molecular motor based on photochemical/thermal cycling has not yet been experimentally reduced to practice.

A number of research groups have designed rotary motors based on catenanes, which contain interlocked rings. In this context, a review by Kay and Leigh (*6*) describes physical considerations for molecular devices, including switches and motors. Leigh, Stoddart, and others have also produced numerous molecular devices based on rotaxanes, in which a ring is constrained to travel along a molecular rod between bulky end groups, or stoppers (*3-6*). In rotaxane switches, a ring is shuttled between different stations by means of light or electron-transfer signals. Similar propulsion of rings in catenanes can be used to fashion rotary molecular motors.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

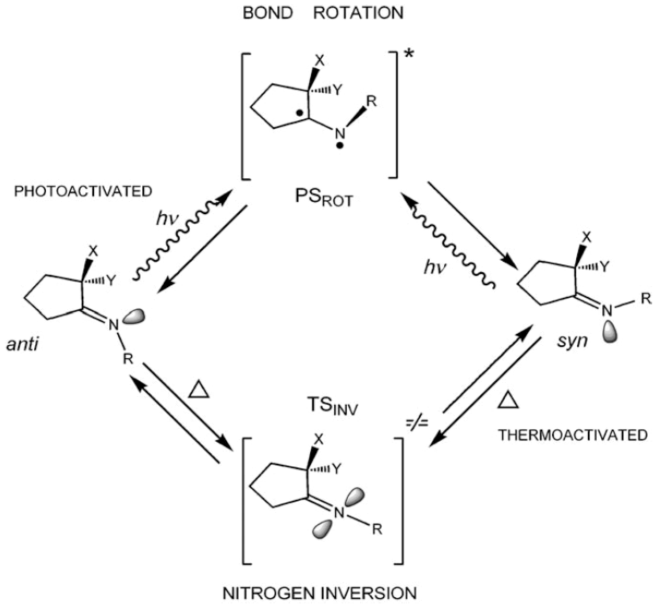


Figure 4. Lehn's putative light-powered chiral imine molecular motor. (Reproduced from reference 25. Copyright 2006 John Wiley & Sons, Inc.).

An Electrically-Driven Molecular Rotor

Rapenne and coworkers have designed an 'electron-triggered motor' containing a wheel, which bears an electroactive group (EG) on each "spoke" (Figure 5) (*26*). Oxidation of the EG closest to the anode causes electrostatic repulsion and motion of the wheel. Adequate insulation between the EGs is required to prevent intramolecular electron transfer. The Rapenne group has synthesized a key target molecule using a cyclopentadienyl ruthenium complex as the central bearing for the wheel. The arms contain [2.2.2]bicyclooctane groups, which have been proven to provide good insulation. The next steps are to anchor the molecule to a surface and address it as a single molecule by means of two metallic nanoelectrodes. This system differs from a ratchet motor, in that it could slip backward when no voltage is applied.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

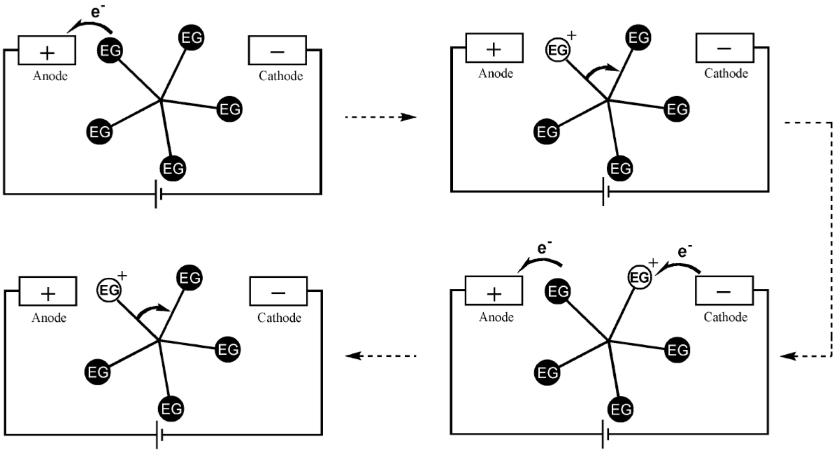

e-
+
Anode
EG
EG+
−
Cathode

Figure 5. Schematic diagram of Rapenne's electrically-driven molecular rotor. (Reproduced from reference 26. Copyright 2008 IUPAC).

A Novel, Photon-Driven Ratchet Motor Design

The authors' own molecular motor design (Figure 6) features a photoactive dibenzofulvene rotor, which is sterically geared into a chiral triarylmethane stator. Light absorption twists the double bond about torsional "drive angle" θ. Steric gearing causes a correlation between the directions of rotation about θ and the "axle angle", Φ. The tilt of the three benzene rings in the stator causes the twisting energies to be different for the two possible directions of rotation. Molecular dynamics calculations have shown that the enantiomer displayed in Figure 6 ($Z = CH_2C{\equiv}C$) should have an approximately 80% preference for counterclockwise vs. clockwise rotation of the rotor with respect to the stator (*27*).

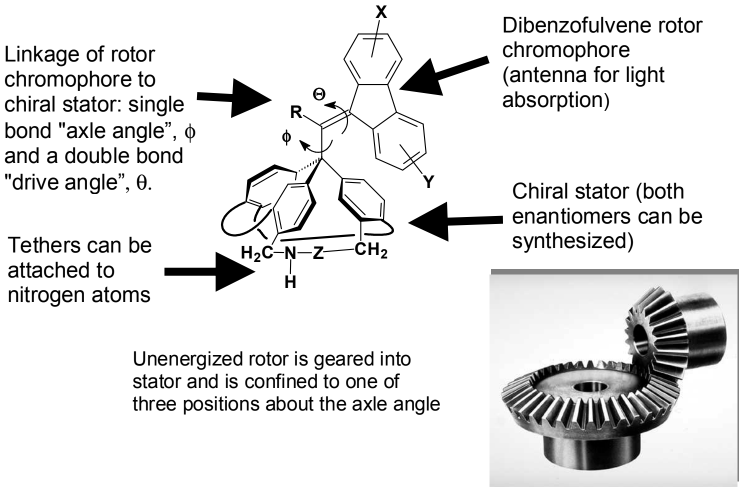

Linkage of rotor chromophore to chiral stator: single bond "axle angle", ϕ and a double bond "drive angle", θ.
Tethers can be attached to nitrogen atoms
X
Y
R
Θ
ϕ
H₂C–N–Z–CH₂
H
Dibenzofulvene rotor chromophore (antenna for light absorption)
Chiral stator (both enantiomers can be synthesized)
Unenergized rotor is geared into stator and is confined to one of three positions about the axle angle

Figure 6. The authors' design of a photon-driven ratchet molecular motor.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

As shown by the retrosynthetic disconnection in Figure 7, the target motor may be synthesized, in principle, by reaction of the metalated stator (e.g. M = Li) with a 9-halomethylenefluorene rotor (e.g. X = Br). This convergent approach involves separate syntheses of the chiral stator and various rotors bearing substituents that can be used to fine tune the performance of the motor. Hence, various unsymmetrically substituted 9-bromomethylenefluorenes were synthesized in order to test the critical rotor-stator coupling reaction, and to investigate the efficiency of light-induced twisting about the double bond in rotor model compounds. The latter factor is very important because the overall efficiency of motor function is limited by the quantum yield of light-induced rotation about the photoactive double bond.

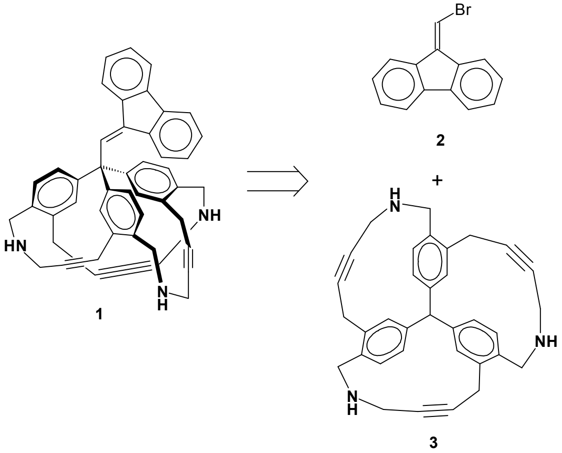


Figure 7. Retrosynthesis of target molecular motor (left), disconnecting the dibenzofulvene rotor (upper right) and the chiral triarylmethane stator (lower right).

Synthesis of the E and Z isomers of a model rotor, 2-*t*-butyl-9-(triphenylethylidene)fluorene, is shown in Figure 8 (*28*). 2-*t*-Butylfluorene was oxidized to the ketone, which was then converted to the bromomethylene derivative by a Wittig reaction. Substitution of bromo by the triphenylmethyl (trityl) group by an addition-elimination reaction with trityllithium furnished the target model compound. The E and Z isomers were separated by chromatography and crystallization for photoisomerization studies. The reactions employed in this synthesis set the stage for preparation of further model compounds, as well as the target molecular motor.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

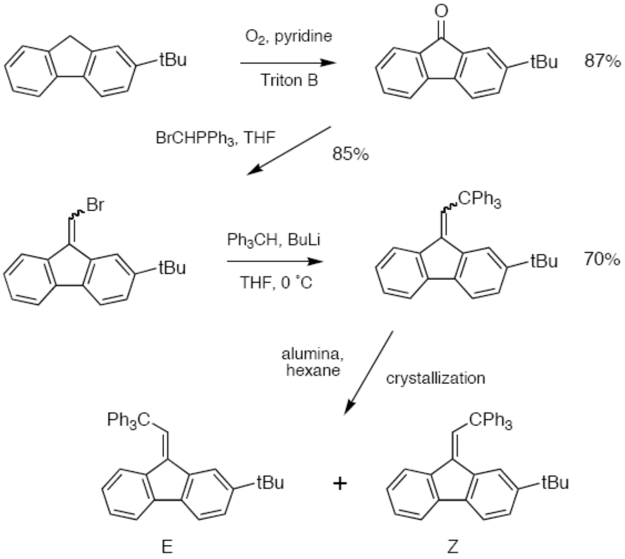

O$_2$, pyridine
Triton B
tBu
87%
BrCHPPh$_3$, THF
85%
Br
Ph$_3$CH, BuLi
THF, 0 °C
CPh$_3$
70%
alumina, hexane
crystallization
Ph$_3$C
+
E
Z

Figure 8. Synthesis of (E)- and (Z)-2-tert-butyl-9-(2,2,2-triphenylethylidene)fluorenes (28).

Model rotors with nitrogen-containing substituents were also synthesized (Figure 9). Here, the 2,5-dimethylpyrrolidin-1-yl group was used to protect an amino group in the 2-position of fluorene. Thus, three model compounds, containing 2,5-dimethylpyrrolidin-1-yl, amino, or nitro groups, were prepared. Attempts to make the 2-nitro rotor model by a shorter route failed, due to the low solubility of 2-nitrofluorenone in most solvents.

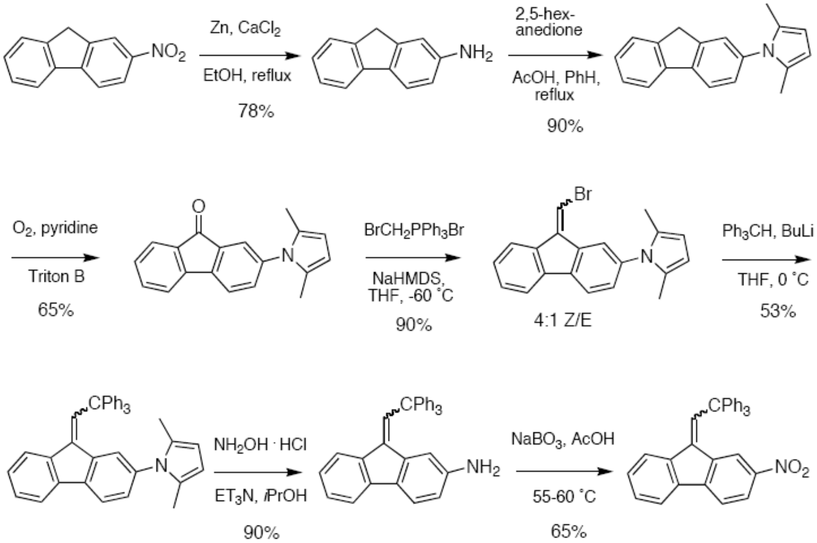

NO$_2$
Zn, CaCl$_2$
EtOH, reflux
78%
NH$_2$
2,5-hex-anedione
AcOH, PhH, reflux
90%
N
O$_2$, pyridine
Triton B
65%
O
BrCH$_2$PPh$_3$Br
NaHMDS, THF, -60 °C
90%
Br
4:1 Z/E
Ph$_3$CH, BuLi
THF, 0 °C
53%
CPh$_3$
NH$_2$OH · HCl
ET$_3$N, *i*PrOH
90%
NaBO$_3$, AcOH
55-60 °C
65%

Figure 9. Synthesis of model rotors with nitrogen substituents, especially 2-nitro-9-(2,2,2-triphenylethylidene)fluorene.

Of the four rotor models synthesized, only the 2-*t*-butyl and 2-nitro analogs underwent photoisomerization. The quantum yield (ratio of isomerization to light absorption) was measured at various excitation wavelengths (λ_{ex}), both for E→Z (Φ_{EZ}) and Z→E (Φ_{ZE}) isomerization. The results are shown in Table I along with the mole fractions of the E and Z isomers at the photostationary state ($([E]/[Z])_{ps}$. The quantum yields for the *t*-butyl rotor model were uniformly less than 0.1, while the quantum efficiencies of photoisomerization in the nitro analog were as high as 0.26. In both cases, excitation of the longest wavelength electronic transition gave the highest quantum yield. These results demonstrate photoactivity of the dibenzofulvene chromophore for the first time and show adequate quantum efficiency for practical operation of the designed molecular motor.

So far, our synthetic efforts have not yielded substantial quantities of the target, chiral, triarylmethane stator. Current efforts are aimed at the synthesis of other C_3-symmetric polycyclic skeletons that also show promise as chiral stators, according to molecular dynamics simulations.

Table I. Photoisomerization Quantum Efficiency for 2-Substituted 9-(2,2,2-triphenylethylidene)fluorenes

Substituent	λ_{ex} *(nm)*	Φ_{EZ}	Φ_{ZE}	$([E]/[Z])_{ps}$
t-Butyl	266	0.09	0.08	0.43/0.57
t-Butyl	280	0.04	0.04	0.47/0.53
t-Butyl	320	0.09	0.07	0.42/0.58
Nitro	266	0.25	0.17	0.38/0.62
Nitro	290	0.26	0.17	0.39/0.61
Nitro	355	0.21	0.12	0.33/0.67

Conclusions

We have seen that biomotors driven by proton gradients or by ATP hydrolysis have evolved to perform many functions. Artificial motors designed to be driven by chemical reactions, photon absorption, or electron transfer have been created by *Homo sapiens*. The authors of this chapter have designed a photon-driven ratchet motor consisting of a dibenzofulvene rotor and a chiral, triarylmethane stator. Quantum yields for substituted dibenzofulvene rotors have been measured and are adequate for practical operation of the target motor.

Acknowledgements

The authors express their gratitude for the many contributions of their collaborators, including Prof. Christine Cremo, Prof. John Frederick, James Barr, Kelly Burtt, Stephanie Everhart, Dr. Nicholas Hext, HyunJong Kim, Daniel Phillips, Rolando Procupez, and Dr. Wayne Stanbery. Support of the authors' work by the US National Science Foundation via Nanoscale

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

Interdisciplinary Research Team (NIRT) grant # CHE 0210549, is also gratefully acknowledged.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

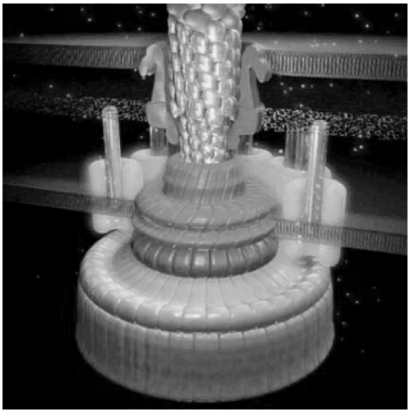


Plate 1.(see color insert 3) Representation of the bacterial flagellar motor. (Reproduced with permission from reference 12.)

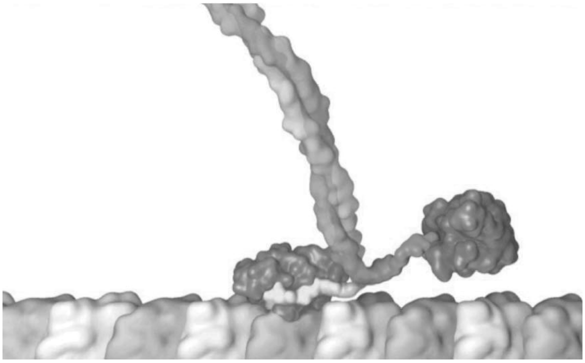


Plate 2.(see color insert 4) Representation of kinesin, a two-headed processive motor, which attaches to and "walks" along a microtubule, shown at bottom. (Partially reproduced from reference 15. Copyright 2000 American Association for the Advancement of Science.)

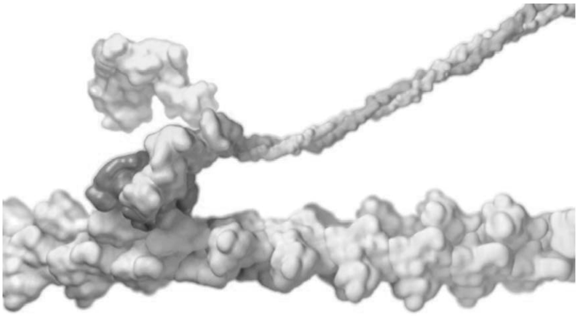


Plate 3.(see color insert 4) Representation of muscle myosin, a two-headed nonprocessive motor, which attaches to actin and delivers one stroke before detaching. (Partially reproduced from reference 15. Copyright 2000 American Association for the Advancement of Science.)

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

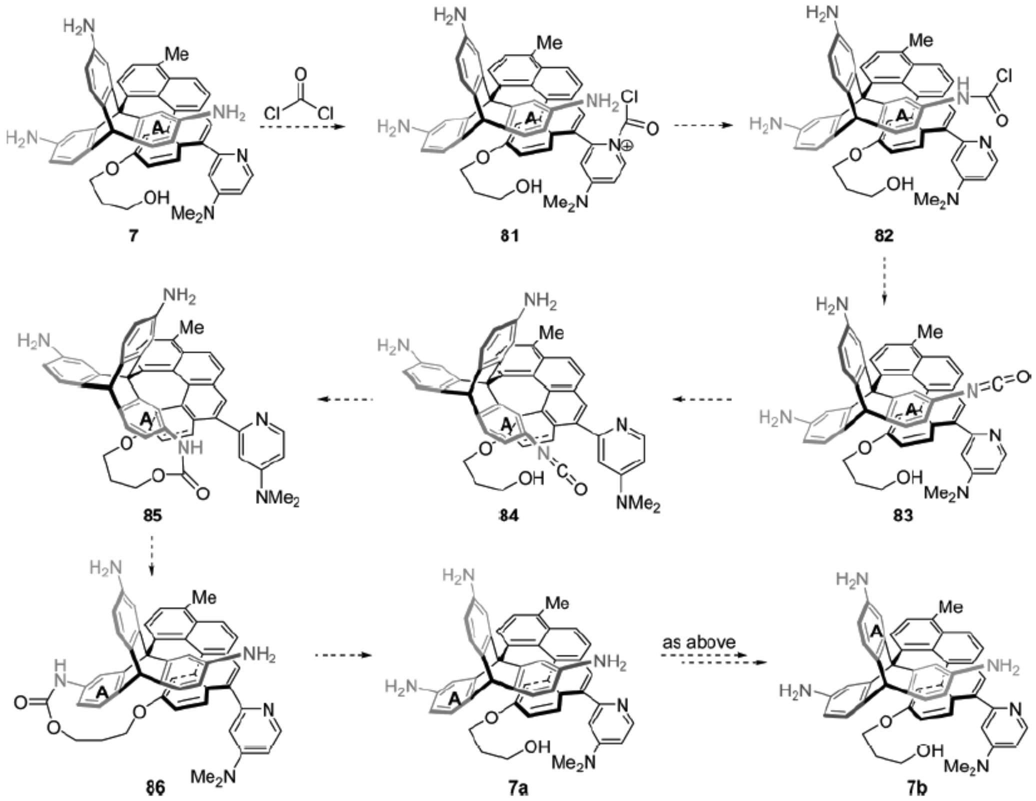

7
81
82
83
84
85
86
7a
7b
as above
A
Me
H_2N
NH_2
OH
Me_2N
NMe_2
N
O
Cl
NH
N=C=O

Plate 4. (see color insert 5) Kelly's chemically-powered ratchet-type rotary motor. (Reproduced from reference 20. Copyright 2007 American Chemical Society).

References

1. Bell, T. W.; Hext, N. M. *Chem. Soc. Rev.* **2004,** *33,* 589–598.
2. *Optical Switches and Sensors;* Ramamurthy, V.; Schanze, K. S., Eds.; Marcel Dekker: New York, 2001.
3. Balzani, V.; Credi, A.; Raymo, F. M.; Stoddart, J. F. *Angew. Chem. Int. Ed. Engl.* **2000,** *39,* 3348–3391.
4. *Structure & Bonding Vol. 99: Molecular Machines and Motors;* Sauvage, J.-P.; Ed.; Springer-Verlag: Berlin, Germany, 2001.
5. Balzani, V.; Venturi, M.; Credi, A. *Molecular Devices and Machines: A Journey into the Nanoworld;* Wiley-VCH: Weinheim, 2003.
6. Kay, E. R.; Leigh, D. A. *Pure Appl. Chem.* **2008,** *80,* 17–29.
7. van den Heuvel, M. G. L.; Dekker, C. *Science (Washington, D.C.)* **2007,** *317,* 333–336.
8. Elston, T.; Wang, H.; Oster, G. *Nature* **1998,** *391,* 510–513; for animation, see: URL http://nature.berkeley.edu/~hongwang/Project/ATP_synthase/.
9. Noji, H.; Yasuda, R.; Yoshida, M.; Kinosita, K., Jr. *Nature* **1997,** *386*, 299–302; for animations, see: URL http://www.res.titech.ac.jp/~seibutu/.
10. Soong, R. K.; Bachand, G. D.; Neves, H. P.; Olkhovets, A. G.; Craighead, H. G.; Montemagno, C. D. *Science (Washington, D.C.)* **2000,** *290,* 1555–1558.
11. Berg, H. C. *Ann. Rev. Biochem.* **2003,** *72,* 19–54.
12. Courtesy of Prof. Keiichi Namba, Nanobiology Laboratories Protonic Nanomachine Group; for animations, see: URL http://www.fbs.osaka-u.ac.jp/eng/labo/09a.html.
13. Behe, M. J. *Darwin's Black Box: The Biochemical Challenge to Evolution;* The Free Press: New York, 1996.
14. Miller, K. R., URL http://millerandlevine.com/km/evol/design2/article.html.
15. Vale, R. D.; Milligan, R. A. *Science (Washington, D.C.)* **2000,** *288,* 88–95.
16. Amos, L. A. *Cell. Mol. Life Sci.* **2008,** *65,* 509–515.
17. Stills from animations created by Graham Johnson of www.fivth.com and The Scripps Research Institute (TSRI) for Ron Milligan of TSRI and Prof. Ron Vale, UCSF; for animations, see: URL http://www.scripps.edu/cb/milligan/projects.html, or www.fivth.com.
18. Feringa, B. L.; van Delden, R. A.; Koumura, N.; Geertsma, E. M. *Chem. Rev.* **2000,** *100,* 1789–1816.
19. Kelly, T. R.; Silva, R. A.; De Silva, H.; Jasmin, S.; Zhao, Y. *J. Am. Chem. Soc.* **2000,** *122,* 6935–6949.
20. Kelly, T. R.; Cai, X.; Damkaci, F.; Panicker, S. B.; Tu, B.; Bushell, S. M.; Cornella, I.; Piggott, M. J.; Salives, R.; Cavero, M.; Zhao, Y.; Jasmin, S. *J. Am. Chem. Soc.* **2007,** *129,* 376–386.
21. Dahl, B.; Branchaud, B. P. *Org. Lett.* **2006,** *8,* 5841–5844.
22. Fletcher, S.P.; Dumur, F.; Pollard, M.M.; Feringa, B.L. *Science* (*Washington, D.C.*) **2005**, *310*, 80–82.
23. Credi, A. *Aust. J. Chem.* **2006,** *59,* 157–169.
24. Pollard, M. M.; Meetsma, A.; Feringa, B. L. *Org. Biomol. Chem.* **2008,** *6,* 507–512.
25. Lehn, J.-M. *Chem. Eur. J.* **2006,** *12,* 5910–5915.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

26. Jacquot de Rouville, H.-P.; Vives, G.; Rapenne, G. *Pure Appl. Chem.* **2008,** *80,* 659–667.
27. Kelly Burtt, *unpublished results*.
28. Barr, J. W.; Bell, T. W.; Catalano, V. J.; Cline, J. I.; Phillips, D. J.; Procupez, R. *J. Phys. Chem.* **2005,** *A109,* 11650–11654.

Downloaded by DUKE UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch012

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

Chapter 13

Coordination-Driven Self-Assembly

S. Russell Seidel and Lori Zaikowski

Department of Chemistry and Physics
Dowling College, Oakdale, NY 11769

This chapter focuses on the field of coordination-driven self-assembly as it pertains to discrete, supramolecular rings and cages. The evolution of the field is examined through relevant historical examples as well as those that elucidate interesting aspects of the self-assembly process or its potential applications. Noteworthy current trends in the field are also explored. Included in this review are supramolecular squares, triangles, pentagons, hexagons, dodecahedra, octahedra, tetrahedra, and other rings / cages.

Introduction

Self-assembly is utilized by nature in a variety of its systems. Such instances are the result of the evolutionary mechanisms that have been shaping the biosphere of our world since the inception of life on Earth. Amongst these natural assembly mechanisms are the self-organization of highly-symmetrical viral coats (*1,2,3*) and the actions of actin filaments in the cytoplasm of our cells (*3,4*).

In their endeavors to mimic biological self-assembly, one of the more prominent avenues that researchers have explored in synthetic self-assembly has been that of employing a coordination-driven motif (*3,5-11*). The logic behind this design strategy relies upon utilizing coordinate covalent (or dative) ligand-metal bonds, which are generally weaker than nonmetal-nonmetal covalent bonds yet stronger and more-directional than the myriad of non-bonding forces that one finds in biologically self-assembled systems. In taking this "middle ground" of sorts, researchers have been able to take advantage of the directionality / spatial-predictability of ligand-metal bonds, while still allowing for a phenomenon known as "self-correction" to occur due to the relative

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

weakness of such bonds (*12*). The notion of "self-correction" relies upon the ligand-metal bonds being weak enough to continually break and reform until the most thermodynamically-favorable product results. This product is presumably, and has often been found to be, the "targeted" two-dimensional macrocycle or three-dimensional cage.

Along these lines, another crucial aspect of the coordination-driven self-assembly strategy has been the often-employed, predictable-product approach. Utilizing complementary subunits with two or more reactive sites apiece, as well as exploiting the angles between these sites, researchers have synthesized a library of pre-determined supramolecular cycles and cages (*3,5-11*). In such instances, the term "complementary" indicates the acceptor (metal-based) / donor (ligand-based) relationship between the subunits, while the number of reactive sites for a given subunit is often described in terms of its "topicity" (i.e., a ditopic donor subunit will typically possess two Lewis basic sites, each capable of forming a dative bond with a metal.) In this modular approach to assembly, the angular / structural information pre-encoded into the building blocks, as well as the ratio in which they are mixed, predetermines the resultant cage or macrocycle in terms of identity, shape, size, and potential functionality. As a consequence of this approach, it is clear that a certain level of rigidity must be possessed by the subunits in order for them to succeed in their mission of pre-determining supramolecular structure. If the linkers are too flexible, a different structure—or mixture of structures—may result. It should be noted, however, that much work has also been done with more flexible linking units, as well (*5*). Such systems tend to be less predictable.

Coordination-driven self-assembly has been the focus of a great deal of research over the past ~15-20 years (*3,5-11*), and it has thus resulted in a large number of self-assembled supramolecular species, with the number of reported products growing steadily and strongly. Amongst these assemblies have been a myriad of two-dimensional rings and three-dimensional cages of all sorts. In this chapter, specific examples that are historically-relevant, particularly illustrative of the field and its diversity, and / or conceptually or functionally significant will be discussed in order to provide a broad overview of the topic.

Two-Dimensional Assembly

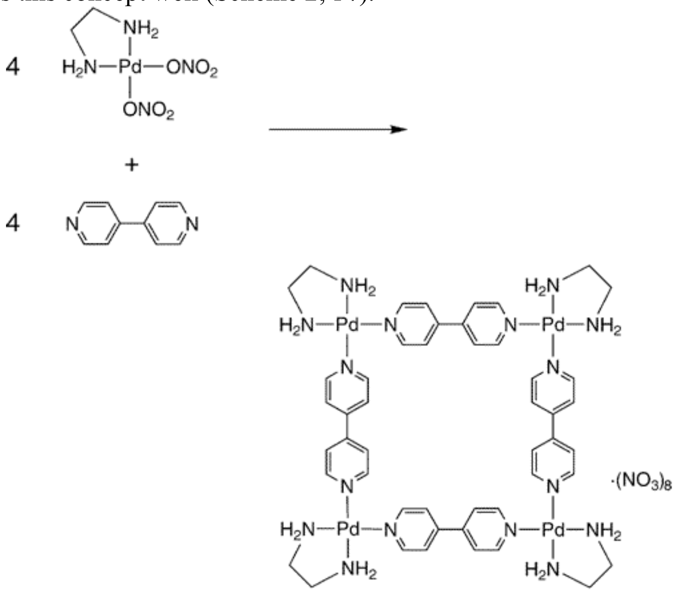
As a means of demonstrating the principles of coordination-driven self-assembly, supramolecular squares are quite useful. Indeed, such macrocycles were amongst the first to be realized due to their conceptual simplicity, and a plethora of such species with a wide variety of subunits and functionality has been reported as the field has developed (*5-8*). There are two primary modes for assembling a supramolecular square: a "four-by-four" methodology and a "two-by-two" methodology.

The "four-by-four" methodology in the assembly of supramolecular squares is quite common, particularly amongst early examples of such systems. This approach is well illustrated by the first transition metal-based supramolecular square, produced by Fujita and co-workers in 1990 (*8,13*). Scheme 1 demonstrates the formation of this square from four palladium(II) ditopic, *cis*-

acceptor subunits and four 4,4'-dipyridine ditopic, donor subunits in a water-alcohol solvent system.

This reaction proceeds quantitatively to the predicted product, as is often the case for coordination-driven self-assembly processes. Moreover, it overtly elucidates the general principles of the design motif, wherein the complementary subunits are mixed in the appropriate ratio relative to one another, and these subunits pre-determine the angles, dimensions, and shape of the resultant product. In this case, a square was the desired supramolecular shape, so four ~90° ditopic acceptor units based on square planar palladium(II) were reacted with four linear ditopic donor linkers, the 4,4'-dipyridine ligands. The palladium(II) species formed the corners of the square, while the 4,4'-dipyridine units acted as the sides. The product's identity and properties were predicted by the precursors that were utilized.

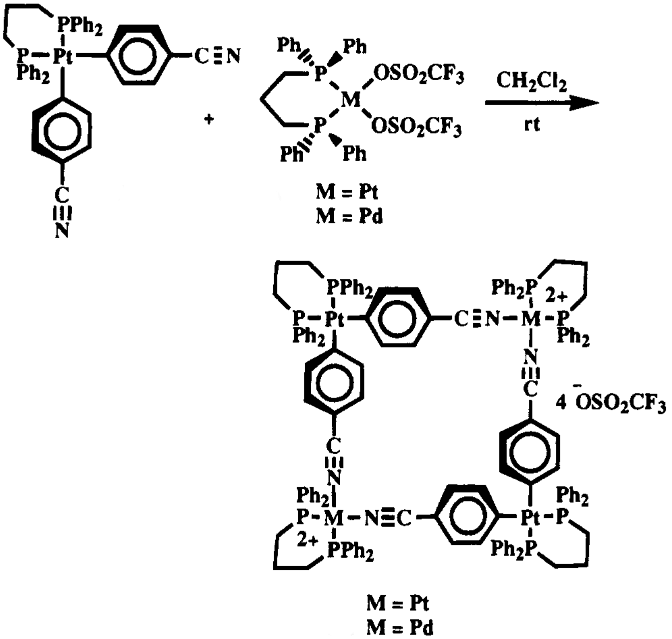
Another route to achieving supramolecular squares is one that is less commonly envisioned, yet is indicative of the diversity found in the field. This is the "two-by-two" methodology, and it was demonstrated well by the work of Stang and co-workers in the mid-to-late 1990's (*14-16*). In this work, both the ditopic acceptor and the ditopic donor subunits were based on square planar transition metal complexes; the donors incorporated platinum(II) and the acceptors incorporated either palladium(II) or platinum(II). Again, the reactive sites about the metal were *cis*- to one another in these tectons, yielding an angle of ~90° between them. With both building blocks being angular, however, the need for a linear linking unit was eliminated, as each resultant square consisted of two donor corners and two acceptor corners. The assembly of the earliest set of such squares, undertaken in methylene chloride at room temperature, illustrates this concept well (Scheme 2; *14*).



Scheme 1. First transition metal-based supramolecular square: four-by-four methodology. Reproduced from reference 8. Copyright 2005 American Chemical Society.

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

It is of interest to note that the design of these systems allowed for the synthesis of mixed-metal squares, such as the platinum(II) – palladium(II) square seen in Scheme 2, as well as others of a similar nature (*15,16*). Utilizing this motif, a variety of such species were prepared, including squares based on cyano-metal and pyridine-metal coordination, squares with differing terminal phosphine ligands on the metal acceptor subunit, and even squares capable of complex host-guest chemistry (*14-16*).



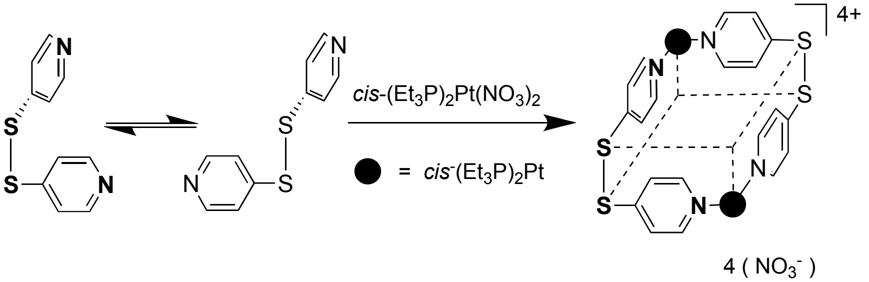
Scheme 2. Two-by-two methodology to form supramolecular squares. Adapted from reference 14. Copyright 1994 American Chemical Society.

In 2001, an interesting variation on the "two-by-two" square system was also reported by Stang and co-workers (*17*). In this work, a less-definite linking unit, 4,4'-dithiodipyridine, was used in conjunction with a ditopic, *cis*-acceptor subunit based on square planar platinum(II) to give a "staggered" square (Scheme 3).

While the metal acceptor complex had reaction sites ~90° apart as expected, so, in a sense, did the ditopic donor, 4,4'-dithiodipyridine. It was known that 4,4'-dithiodipyridine, in solution, exists in two enantiomeric forms that are in rapid equilibrium with one another via rotation about the S-S single bond. Further, it was known that the C-S-S-C dihedral angle in the two enantiomers of this subunit was ~90°. These facts put the nitrogen atoms of the pyridine groups in the molecule ~90° apart (torsion angle-wise) when either enantiomer was viewed down the S-S bond axis. Stang and co-workers made use of this conformationally-defined torsion angle as a means of encoding angular information into the resultant, "staggered" supramolecular square. As was determined by X-ray diffraction data, the donor linker indeed maintained a C-S-S-C dihedral angle of ~90° in the final assembly. This was one of the first

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

instances in which a conformationally-defined angle was utilized as the basis for a pre-determined assembly of this type.



Scheme 3. Two-by-two "staggered" square. Adapted from reference 17. Copyright 2001 American Chemical Society.

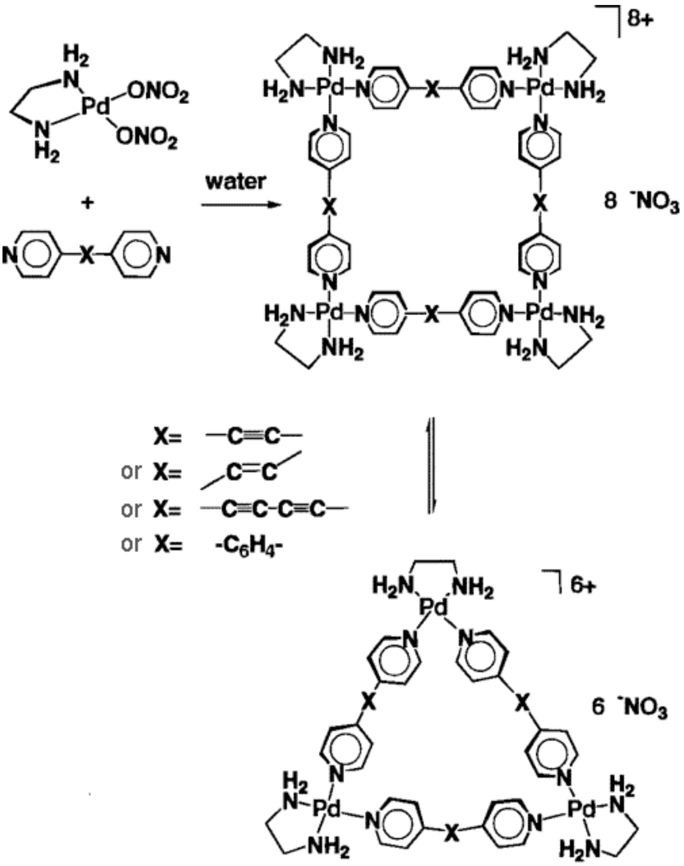
Under certain circumstances, particularly those involving more flexible subunits, an equilibrium is found between square and triangular supramolecular species (*5,18-20*). Work by Fujita and co-workers serves to illustrate such systems well (*5,18*). In this research, square-triangle equilibria are established in which the resultant assemblies are based upon the ditopic, *cis*-acceptor, palladium(II) subunit commonly employed by the Fujita group along with a variety of flexible, 4,4'-dipyridyl-based linking units (Scheme 4).

The cause of the equilibria between the two different macrocycles in each case has been attributed to the increased flexibility of the organic linker and its effects on the thermodynamics of the system (*5,18*). The common logic in the field is that entropic considerations will generally favor the formation of the smallest possible assembly, as this should lead to a greater number of product molecules in solution from given quantities of starting materials. Opposing such entropic factors are enthalpic considerations based on minimizing angular strain in the system. When the tectons making up the assembly are more rigid, the latter often wins out, and only a single, predictable product results (i.e., a square). In the case of the square-triangle equilibria presented in Scheme 4, the reduced rigidity of the "linear" subunits allows for the entropically-favored triangle to better compete with the enthalpically-favored square, and an equilibrium between the two is established. In all fairness, however, such an argument may well be oversimplified, as such factors as solvent effects, ion-pairs, and the like are not fully considered.

Regardless of its source, there is indeed an apparent drive toward smaller assemblies. This is the case even in some instances involving seemingly rigid subunits where ring strain should be the over-riding factor. Again, supramolecular triangles provide an interesting glimpse into this phenomenon (*21-23*).

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013



Scheme 4. Equilibrium between square and triangular supramolecular species. Adapted from reference 5. Copyright 2000 American Chemical Society.

To illustrate, Cotton and co-workers were able to synthesize a triangle based on a di-rhodium ~90° acceptor subunit and oxalate bridging units (*21*). The di-rhodium complexes served as the vertices of the triangle, while the oxalates served as ditopic, chelating sides (Figure 1).

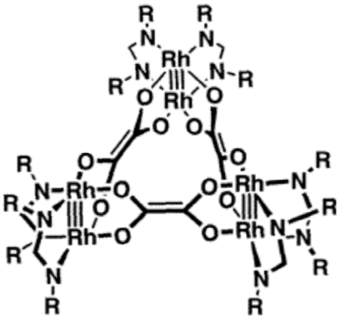


Figure 1. Supramolecular triangle. R = p-MeOC$_6$H$_4$. Adapted from reference 5. Copyright 2000 American Chemical Society.

According to their report, this structure resulted originally from attempts to synthesize an analogous square structure. The square structure (Figure 2) was ultimately realized by a relatively minor modification of the synthetic procedures (a change in the relative ratios of the precursors).

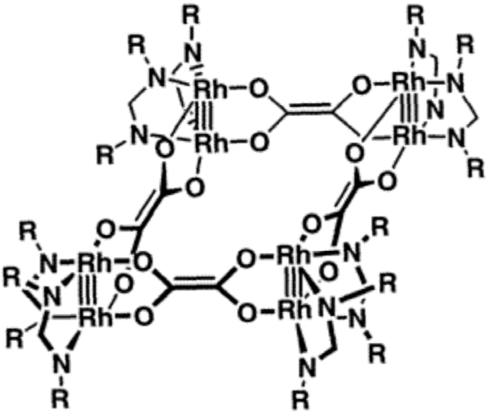
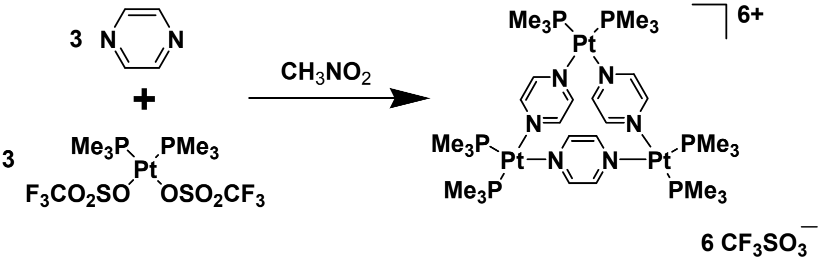


Figure 2. Supramolecular square. R = p-MeOC$_6$H$_4$. Adapted from reference 5. Copyright 2000 American Chemical Society.

Indeed, logic would tend to dictate that the square structure would be the sole product of this type of reaction and that a triangle would not be formed. This should be particularly true given the rigid nature of the oxalate linkers. Yet, under the initial reaction conditions, a triangle was the isolated species. In this triangle, the oxalates were found to be "bowed" and thus compensated for the potential angular strain in the system (*21*). Cotton and co-workers were initially unable to report the cause of this apparent deviation from product prediction and chose not to hazard a guess. In a later report (*22*), they revisited these triangle and square assemblies and noted that, in solution, an equilibrium between the two was possible, and that the existence of the two species could be justified based on entropy and enthalpy considerations.

Another unexpected triangle with rigid and small subunits was reported by Stang and co-workers (*23*). In this work, pyrazine, a ditopic donor, was reacted with a ditopic, *cis*-acceptor, square-planar platinum(II) complex (Scheme 5).



Scheme 5. Unexpected triangular species with rigid and small subunits.

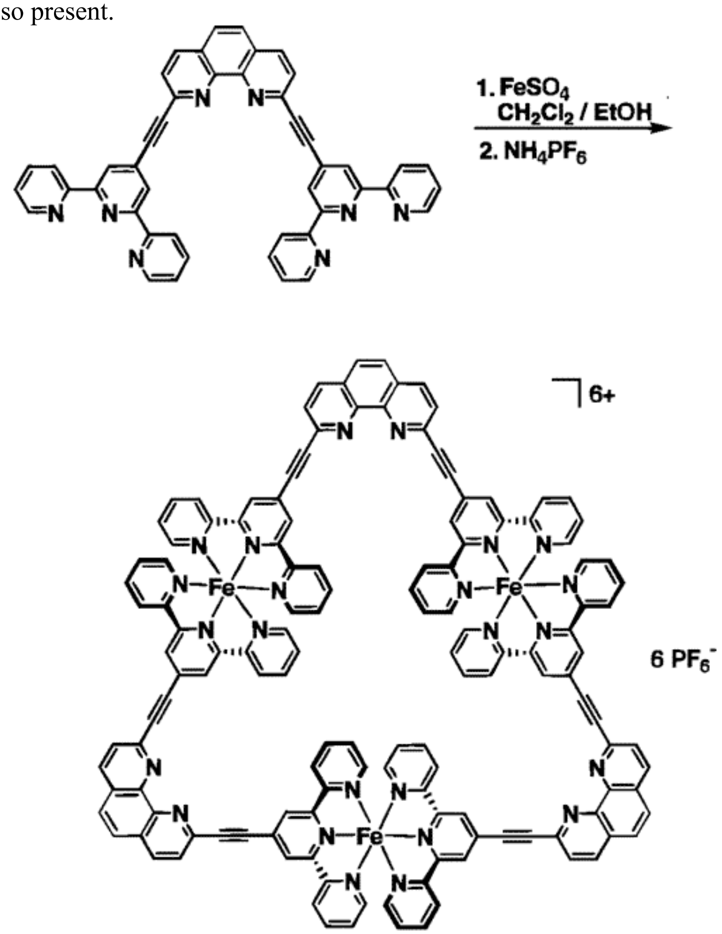
The formation and isolation of a triangular species from such small and inflexible subunits, especially considering that the organic sides of the triangle

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

were the relatively tiny pyrazine donors, was quite unexpected. Indeed, the angular strain in such a system should be quite large, yet the smaller triangle was still the isolated product.

Not all triangular assemblies are the result of either inexplicable phenomena and / or systems that would be solely squares but for flexible linking units. In fact, pre-designed triangles have also been reported, although they are relatively rare. One such report was made by Ziessel and co-workers, in which three iron(II) centers were reacted with and coordinated by three *bis*-terpyridine-based, ditopic, chelating ~60° donor subunits (Scheme 6; *5,24*).

This interesting take on coordination-driven self-assembly relied on the fact that each iron(II) cation was six-coordinate and thus chelated by two different tridentate terpyridyl units of two different linkers. The iron(II) centers thus served as the "glue" in holding together the triangle-defining organic tectons. It should be noted that evidence for a square-like species in the reaction mixture was also present.

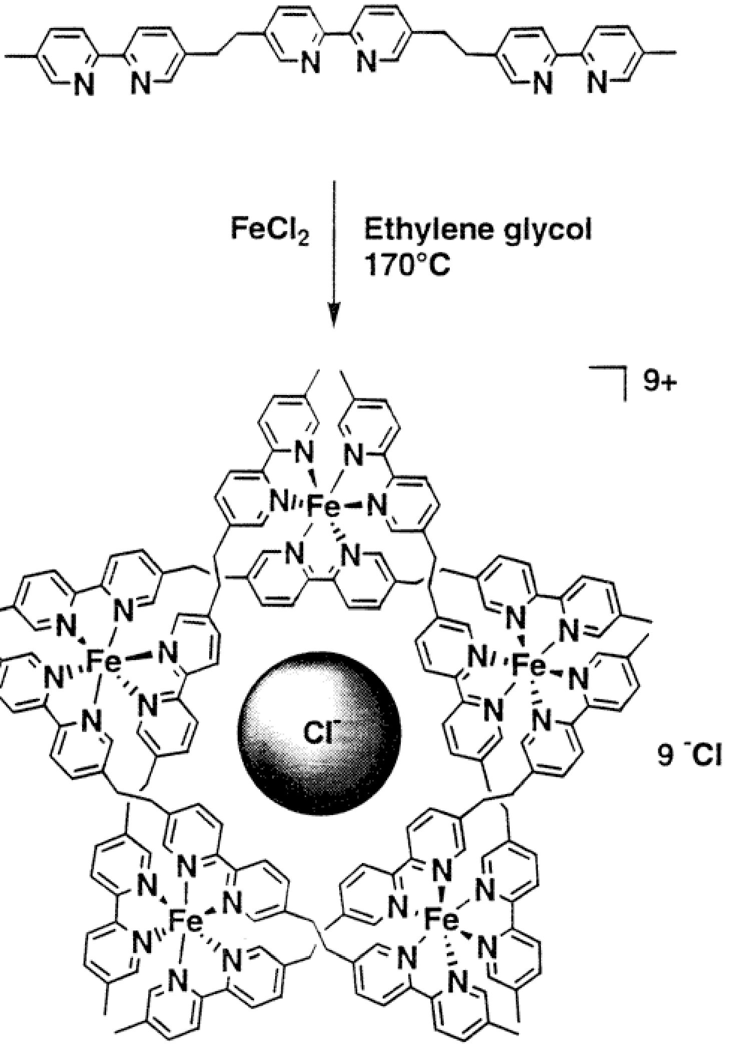


Scheme 6. Pre-designed triangle. Reproduced from reference 5. Copyright 2000 American Chemical Society.

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

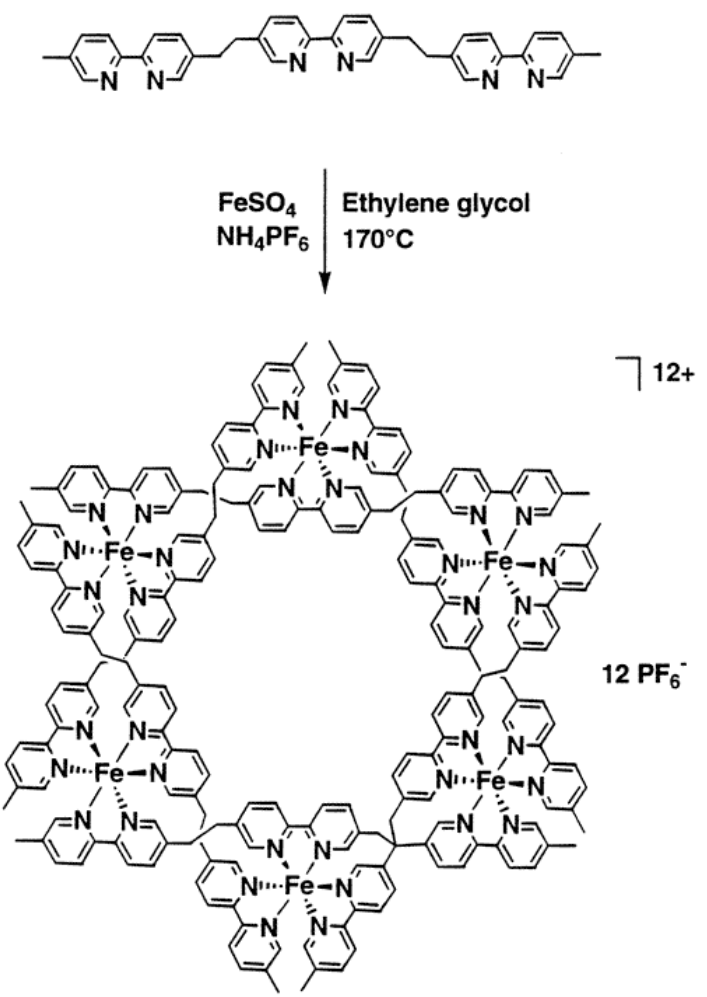
Aside from supramolecular squares and triangles, a variety of other two-dimensional species have been synthesized over the past number of years. Included amongst these are pentagons and hexagons (Schemes 7 and 8).

In terms of pentagons and hexagons, the work of Lehn and co-workers stands out as being amongst the first syntheses of such systems (*5,25,26*). In a manner similar to that seen in the work of Ziessel, six-coordinate iron(II) centers



Scheme 7. Supramolecular pentagon. Reproduced from reference 5. Copyright 2000 American Chemical Society.

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013



Scheme 8. Supramolecular hexagon. Reproduced from reference 5. Copyright 2000 American Chemical Society.

serve as the "glue" holding together the organic donor linkers, whose mode of attachment is via chelation of three different metals each. In these cases, each iron(II) is chelated in a bidentate fashion by three 2,2'-bipyridine subunits of three different donors. The donor subunits are clearly lacking in terms of rigidity, being more like molecular "strings" than the strictly-defined tectons seen in previous examples in this chapter. While this necessarily reduces the initial predictability of the reactions' outcomes, it also allows for a "tuning" of the system and the formation of more than one possible product, depending upon the starting materials.

Indeed, Lehn and co-workers were able to produce a supramolecular, star-like pentagon when the iron starting material was iron(II) chloride (Scheme 7; *5,26*). Additionally, they were able to produce a supramolecular, star-like

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

hexagon when the iron starting material was iron(II) sulfate and ammonium hexafluorophosphate was utilized as a means of precipitation (Scheme 8; *5,26*).

Lehn and co-workers were able to reasonably demonstrate, through the use of multiple iron(II) starting materials differing only in their counter-anions, that the chloride was likely templating the assembly of the pentagonal species. Likewise, the sulfate was likely templating the assembly of the hexagonal species. From their studies, it was quite evident that this effect was based purely on the dimensions of the anion that was present in the reaction mixture. The pentagon and hexagon produced by Lehn and co-workers fall into the category of compounds known as circular double helicates.

Supramolecular pentagons based upon less flexible subunits and a "five-by-five", angle-side approach are quite rare in the literature. One of the first of such species was synthesized by Dunbar and co-workers in 2001 (*27*). This work involved the reaction of five molar equivalents of $[Ni(CH_3CN)_6][SbF_6]_2$ with five molar equivalents of 3,6-bis(2-pyridyl)1,2,4,5-tetrazine (bptz) to give a pentagonal assembly of the formula $[Ni_5(bptz)_5(CH_3CN)_{10}][SbF_6]_{10}$. Figure 3 shows the structure of the resultant pentagon, as well as the fact that one of the SbF_6^- anions occupies its central cavity as a guest.

In this structure, each of the five bptz donor units bonds to two different nickel(II) centers via bidentate chelation at each metal. Each nickel center has an octahedral-like, six-coordinate geometry, with two different bptz ligands chelating it and two solvent acetonitrile molecules bonded to it in a *cis*-fashion as monodentate ligands. Given that the ideal angles at the vertices of a pentagon should be ~108° and that the angles at the vertices of this structure are ~90°, it might seem odd that a pentagon would result instead of a square. Indeed, in the presence of the correct anion, a smaller anion, such systems have been found to give square assemblies (*28,29*). It is the large SbF_6^- and its templating effects that allowed for the synthesis of the pentagonal species. Any possible angular strain that could be developed in the structure by deviating from ideal angles is largely alleviated via a bending of the organic linking units. A series of further and more detailed studies on this type of system has been published by Dunbar and co-workers in the recent past (*29*).

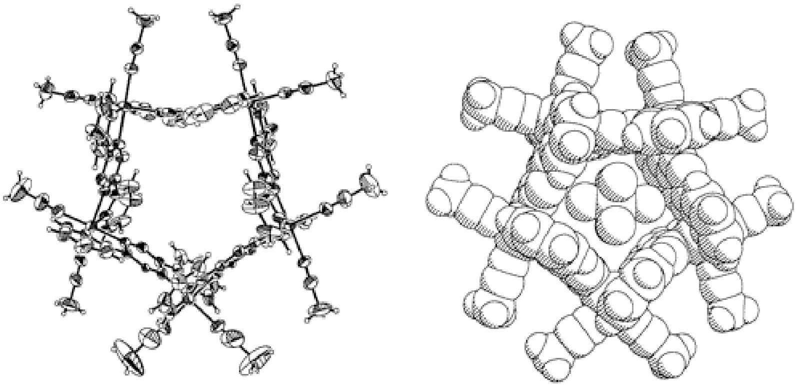


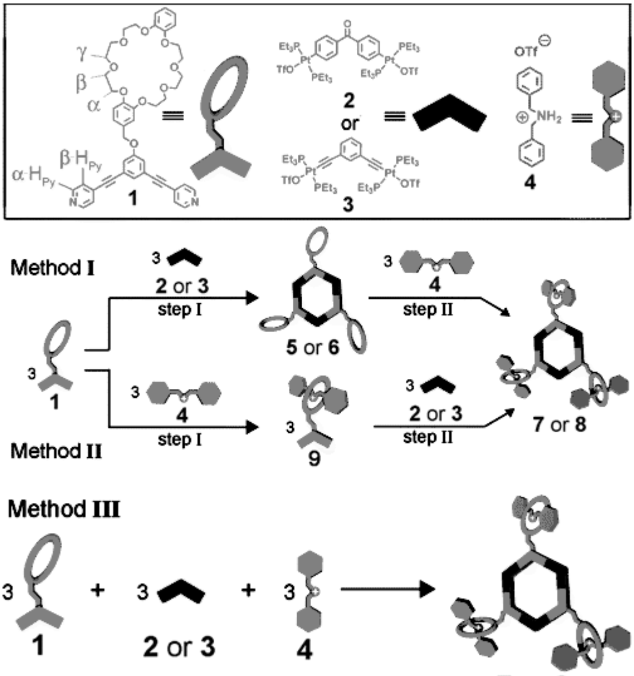
Figure 3. Supramolecular pentagon formed with five-by-five methodology. Reproduced from reference 27. Copyright 2001 American Chemical Society.

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

Supramolecular hexagons via either a “three-by-three” angle-angle approach similar to that of the “two-by-two” squares or a “six-by-six” angle-side approach similar to that of the “four-by-four” squares are much more common than the “five-by-five” pentagons (*5,30-33*). As such, a host of research into these types of systems appears in the literature, and many interesting species abound.

Some of the most recent work of Stang and co-workers gives insight into the wide array of possibilities available for such hexagonal assemblies, and, more generally, for the modular approach to self-assembly itself (*30-33*). One such set of systems is a pair of “three-by-three” angle-angle self-assembled hexagons incorporating three crown-ethers on their peripheries (Scheme 9; *33*).

These hexagonal assemblies each comprise three ~120° ditopic donor corner units datively-bound to three ~120° ditopic platinum(II)-based acceptor corner units. In both hexagons, the same donor linker is utilized, which is also the subunit that possesses the covalently-attached crown-ether substituent. The differences between the two arise solely from the acceptor corner units that are employed. In both assemblies, the host-guest capacity of the peripheral crown-ethers is exploited via the inclusion of a single dibenzylammonium guest in each of the ethers. Significantly, the same two hexagons were each synthesized in three ways: (i) by forming the supramolecular species first and then establishing the host-guest adduct second; or (ii) by forming the host-guest crown ether-dibenzylammonium adduct first and the assembled species second;



Scheme 9. Supramolecular hexagons formed with three-by-three angle-angle methodology. Adapted from reference 33. Copyright 2007 American Chemical Society.

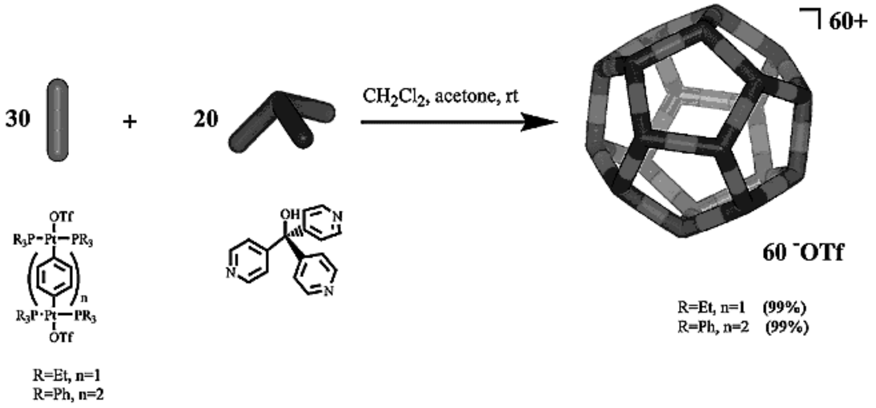
Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

or (iii) in a single preparative step in which all three components are present from the onset. Stang and co-workers have also synthesized similar "three-by-three"angle-angle hexagons and the corresponding "six-by-six" angle-side hexagons with a variety of other peripheral substituents, ranging from attached ferrocene groups (*30*) to an assortment of dendrimeric groups (*31,32*).

Three-Dimensional Assembly

As the field of coordination-driven self-assembly has progressed and its precepts have become better understood over the years, it seems only reasonable that it would begin to transition from two-dimensional rings to three-dimensional cages. Indeed, this has occurred in earnest. To date, the number and variety of such species have grown so large (*3,5,7,8,10,11*), that it is beyond the scope of this work to do full justice to the diversity of the field. That said, certain three-dimensional, self-assembled species stand out as prominent contributors to the development of the field. These will be highlighted, as will some of the more recent research endeavors of note.

Serving as an example of how far the three-dimensional self-assembly motif can go, Stang and co-workers synthesized two of the largest, discrete, self-assembled entities to date: a pair of dodecahedra (*34*). These species were produced via the reaction of twenty equivalents of tris(4-pyridyl)methanol with thirty equivalents of a linear platinum(II)-based, ditopic acceptor (Scheme 10).



Scheme 10. Three-Dimensional self-assembly of dodecahedral species. Adapted from reference 3. Copyright 2002 American Chemical Society.

The tris(4-pyridyl)methanol building blocks, being tritopic ~109.5° donor units, provided the vertices of the dodecahedra as well as the appropriate pre-encoded angular information for the systems. The linear ditopic acceptor subunits, which were different for the two different dodecahedral products, accounted for the sides of the assemblies. As a consequence of the large number of components that came together during the assembly processes, these dodecahedra both reached easily into the nanoscale range, with the smaller of the two having a diameter of ~5.5 nm and the larger a diameter of ~7.5 nm.

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

Indeed, the latter of the two was visible under transmission electron microscopy (Figure 4).

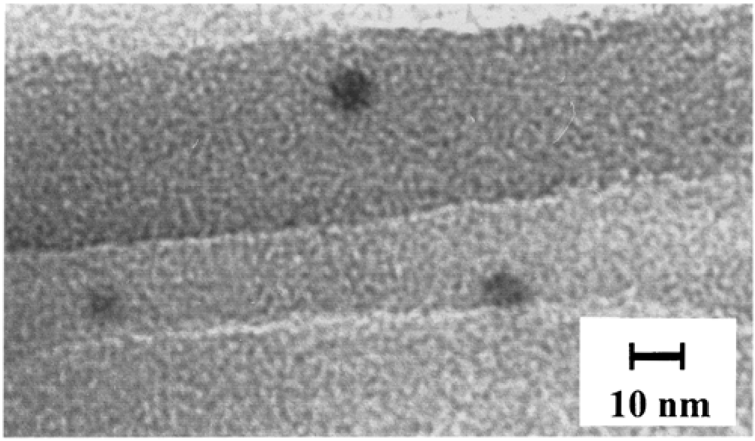


Figure 4. Transmission electron microscopy of dodecadral species with a diameter of ~7.5 nm. Reproduced from reference 34. Copyright 1999 American Chemical Society.

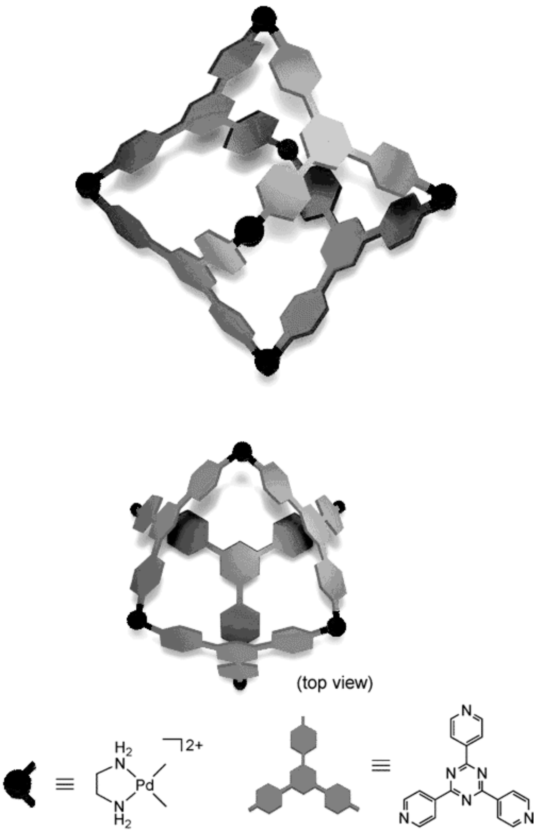
Another crucial set of products in the development of the three-dimensional branch of self-assembly was the octahedral species of Fujita and co-workers (*8,35-41*). The first of such species was the result of the reaction of a palladium(II)-based, ditopic, ~90° acceptor complex and a tritopic, ~120°, planar, donor linker in a 3 : 2 ratio (Scheme 11; *8*).

This architecture was water-soluble and characterized fully, including via X-ray crystallography. The palladium acceptors constituted the vertices of the octahedron, while the tritopic planar linkers spanned every other face. Formally, this species possessed a symmetry that was tetrahedral, as opposed to having a true octahedral symmetry, due to the every-other nature of the faces (*8*).

Significantly, such assemblies were found by Fujita and co-workers to have a number of interesting application-based properties, due largely to the hydrophobic nature of their cavities (*35-41*). Included amongst these uses were a platinum(II)-based analog that served as a "molecular lock" (*37*), host-guest complexes with such octahedra acting as the hosts (*35-37,40,41*), the catalysis / promotion of reactions by these cages (*38,41*), and the isolation of fleeting intermediates in the cavities of such assemblies (*40*).

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013



Scheme 11. Three-dimensional self-assembled octahedral species. Adapted from reference 8. Copyright 2005 American Chemical Society.

Also serving as a prominent step in the growth of three-dimensional assembly was the work of Raymond and co-workers, particularly in the area of synthesizing a variety of supramolecular tetrahedra (*10,11,42-47*). One such assembly is presented in Figure 5 (*42*).

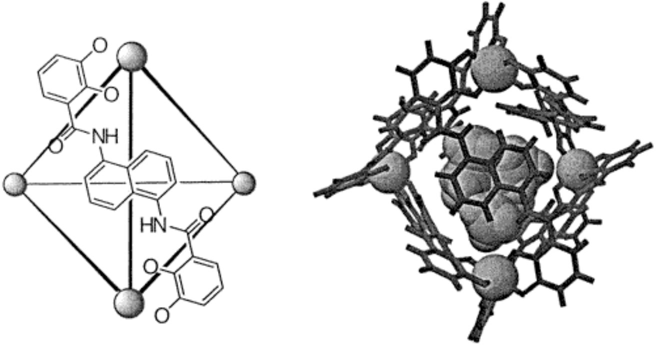


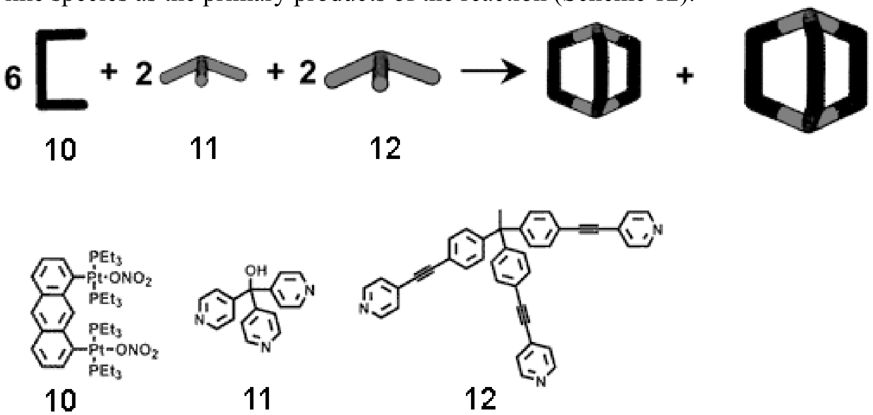
Figure 5. Three-dimensional self-assembled tetrahedral species. Reproduced from reference 42. Copyright 1998 American Chemical Society.

This particular tetrahedron incorporated six-coordinate iron(III) centers as vertices, each chelated in a bidentate fashion by the ends of three different organic linkers. The organic linkers had two bidentate chelating sites apiece, each bonding to a different iron(III) ion in the structure. In all, there were a total of six of such donor units and four of the metal acceptor centers constituting the tetrahedron.

This was but one member of a wide array of similar tetrahedra synthesized and characterized by Raymond and co-workers (*10,11,42-47*). The design motif remained fairly constant between the various systems, and the metals involved were typically either iron(III) or gallium(III). An interesting feature of these assemblies was their ability to act as hosts in host-guest complexes. Indeed, the tetrahedron shown in Figure 5 was found via X-ray structural analysis to possess a tetraethylammonium cation as a guest in its central cavity. Such tetrahedra are currently still under study by Raymond and co-workers and have provided a number of valuable insights into supramolecular phenomena, particularly into the host-guest properties and other unique applications of these entities (*11,44-47*).

Aside from early examples relevant to the evolution of three-dimensional assembly, there have also been a number of more recent advances in the field that are quite noteworthy. Amongst this group are the self-recognizing trigonal prisms of Stang and co-workers (*12*) and the interpenetrating cylinders of Fujita and co-workers (*48*).

Stang and co-workers recently published work in which a ditopic "clip", based upon an anthracene backbone possessing two parallel, platinum(II), monotopic acceptor sites, allowed for a self-recognition phenomenon when mixed simultaneously with two different tritopic, ~109.5° donor subunits in the appropriate ratios (*12*). The structural differences between the two different donor units were primarily founded upon the relative sizes of the two linkers, and their concurrent reaction with the "clip" led to two distinct trigonal prism-like species as the primary products of the reaction (Scheme 12).



Scheme 12. Three-dimensional self-assembly of trigonal prism-like species. Adapted from reference 12. Copyright 2008 American Chemical Society.

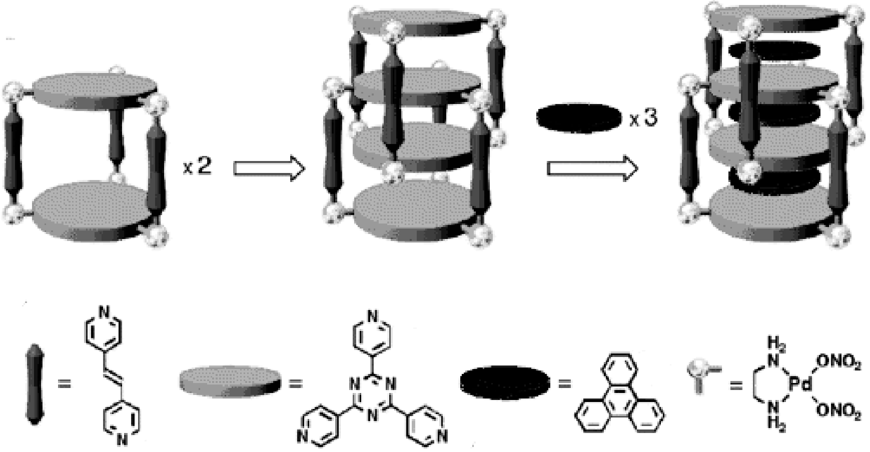
Of the two cages, one consisted exclusively of the "clip" and the smaller of the two donor subunits in a three-to-two ratio, while the other consisted

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

exclusively of the "clip" and the larger of the two donor subunits in the same ratio. It should be noted that the self-recognition phenomenon was not entirely perfect, although close, as there were apparent signs of very minor oligomerization products remaining in the final reaction mixture.

Another recent example of intriguing research into the realm of three-dimensional self-assembly was the work of Fujita and co-workers concerning interpenetrated cylinders (*48*). In contrast to the systems of Stang and co-workers above (*12*), Fujita disallowed a similar self-recognition phenomenon by including templating, aromatic guests in his one-pot syntheses of cylindrical assemblies. Indeed, each cylinder was composed of three linear, ditopic donor subunits clipped to two tritopic, ~120° planar, donor subunits via *cis*-ditopic, palladium(II)-based acceptor clips. The resultant assemblies were all found to be doubly-interpenetrated, and the aromatic molecules served as templating guests within the cages' cavities of interpenetration. A representative example of such a system is shown in Scheme 13.

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013



Scheme 13. Three-dimensional self-assembly of interpenetrated cylinders(note: actual synthesis was necessarily one-pot). Adapted from reference 48. Copyright 2008 American Chemical Society.

The host-guest interactions, and guest-guest interactions where applicable, in such systems were largely the result of π-stacking phenomena, leading to the multiply-stacked-aromatic assemblages. These unique structures were characterized by X-ray crystallography and / or NMR spectroscopy.

Conclusions and Outlook

As it is currently progressing, coordination-driven self-assembly is a truly burgeoning area in the field of chemistry. Indeed, the number of researchers involved in this topic, either directly or indirectly, has seen a tremendous growth since its inception and continues to rise on a yearly basis. As such, this chapter has been formulated in such a manner as to provide a "flavor" of the field in terms of its history, its more prominent developments, and its underlying principles. In so doing, an array of two-dimensional and three-dimensional systems that represent some of the most relevant and important advancements in the area to date has been presented. Where appropriate, potential applications for such systems have also been noted.

It is in this final category, that of applications, that the field has the highest likelihood of substantive, further development. While the number of structures that have been synthesized is vast, the number of true applications for these structures is still relatively small. Given the modular approach of the methodology, as well as the well-established research base that is now in place, it is reasonable to assume that applications will be the next frontier in the area. Examples of host-guest chemistry already abound, and the possibility of "molecular machines" has already been hinted at (*37*). Some examples of reaction promotion and catalysis have appeared in the literature, as well. In many publications, references have been made to the bio-mimicking nature of the paradigm, and it is reasonable to assume that more strides will occur in this direction, too. Indeed, with its inherent advantages over multi-step covalent synthesis and its current popularity, the field of coordination-driven self-assembly will likely be a substantial and productive part of the chemical world of the future.

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

Acknowledgments

This chapter is dedicated to the fond memory of Dr. Robert W. Parry of the University of Utah—Professor Emeritus, former President of the American Chemical Society, Priestly Medal Recipient, and family member, mentor, and friend to so many.

References

1. Horne, R. W. *Virus Structure;* Academic Press: New York, 1974.
2. Cann, A. J. *Principles of Molecular Virology;* Academic Press: San Diego, CA, 1993; pp 1–234.
3. Seidel, S. R.; Stang, P. J. *Acc. Chem. Res.* **2002,** *35,* 972–983.
4. Cooper, J. A. In *Self-Assembling Architecture;* Varner, J. E.; Ed.; Alan R. Liss: New York, 1988; pp 171–177.
5. Leininger, S.; Olenyuk, B.; Stang, P. J. *Chem. Rev.* **2000,** *100,* 853–908.
6. Stang, P. J.; Olenyuk, B. *Acc. Chem. Res.* **1997,** *30,* 502–518.
7. Swiegers, G. F.; Malefetse, T. J. *Chem. Rev.* **2000,** *100,* 3483–3537.
8. Fujita, M.; Tominaga, M.; Hori, A.; Therrien, B. *Acc. Chem. Res.* **2005,** *38,* 371–380.
9. Gianneschi, N. C.; Masar, M. S., III; Mirkin, C. A. *Acc. Chem. Res.* **2005,** *38,* 825–837.
10. Caulder, D. L.; Raymond, K. N. *Acc. Chem. Res.* **1999,** *32,* 975–982.
11. Fiedler, D.; Leung, D. H.; Bergman, R. G.; Raymond, K. N. *Acc. Chem. Res.* **2005,** *38,* 351–360.

12. Zheng, Y.-R.; Yang, H.-B.; Northrop, B. H.; Ghosh, K.; Stang, P. J. *Inorg. Chem.* **2008,** *47,* 4706–4711.
13. Fujita, M.; Yazaki, J.; Ogura, K. *J. Am. Chem. Soc.* **1990,** *112,* 5645–5647.
14. Stang, P. J.; Whiteford, J. A. *Organometallics* **1994,** *13,* 3776–3777.
15. Whiteford, J. A.; Lu, C. V.; Stang, P. J. *J. Am. Chem. Soc.* **1997,** *119,* 2524–2533.
16. Whiteford, J. A.; Stang, P. J. *Inorg. Chem.* **1998,** *37,* 5595–5601.
17. Tabellion, F. M.; Seidel, S. R.; Arif, A. M.; Stang, P. J. *J. Am. Chem. Soc.* **2001,** *123,* 7740–7741.
18. Fujita, M.; Sasaki, O.; Mitsuhashi, T.; Fujita, T.; Yazaki, J.; Yamaguchi, K.; Ogura, K. *J. Chem. Soc. Chem. Commun.* **1996,** 1535–1536.
19. Schweiger, M.; Seidel, S. R.; Arif, A. M.; Stang, P. J. *Inorg. Chem.* **2002,** *41,* 2556–2559.
20. Lee, S. B.; Hwang, S.; Chung, D. S.; Yun, H.; Hong, J.-I. *Tetrahedron Lett.* **1998,** *39,* 873–876.
21. Cotton, F. A.; Daniels, L. M.; Lin, C.; Murillo, C. A. *J. Am. Chem. Soc.* **1999,** *121,* 4538–4539.
22. Cotton, F. A.; Lin, C.; Murillo, C. A. *Inorg. Chem.* **2001,** *40,* 575–577.
23. Schweiger, M.; Seidel, S. R.; Arif, A. M.; Stang, P. J. *Angew. Chem. Int. Ed.* **2001,** *40,* 3467–3469.
24. Romero, F. M.; Ziessel, R.; Dupont-Gervais, A.; Van Dorsselaer, A. *J. Chem. Soc. Chem. Commun.* **1996,** 551–553.
25. Hasenknopf, B.; Lehn, J.-M.; Kneisel, B. O.; Baum, G.; Fenske, D. *Angew. Chem. Int. Ed.* **1996,** *35,* 1838–1840.
26. Hasenknopf, B.; Lehn, J.-M.; Boumediene, N.; Dupont-Gervais, A.; Van Dorsselaer, A.; Kneisel, B.; Fenske, D. *J. Am. Chem. Soc.* **1997,** *119,* 10956–10962.
27. Campos-Fernandez, C. S.; Clerac, R.; Koomen, J. M.; Russell, D. H.; Dunbar, K. R. *J. Am. Chem. Soc.* **2001,** *123,* 773–774.
28. Campos-Fernandez, C. S.; Clerac, R.; Dunbar, K. R. *Angew. Chem. Int. Ed.* **1999,** *38,* 3477–3479.
29. Campos-Fernandez, C. S.; Schottel, B. L.; Chifotides, H. T.; Bera, J. K.; Bacsa, J.; Koomen, J. M.; Russell, D. H.; Dunbar, K. R. *J. Am. Chem. Soc.* **2005,** *127,* 12909–12923.
30. Yang, H.-B.; Ghosh, K.; Zhao, Y.; Northrop, B. H.; Lyndon, M. M.; Muddiman, D. C.; White, H. S.; Stang, P. J. *J. Am. Chem. Soc.* **2008,** *130,* 839–841.
31. Yang, H.-B.; Das, N.; Huang, F.; Hawkridge, A. M.; Muddiman, D. C.; Stang, P. J. *J. Am. Chem. Soc.* **2006,** *128,* 10014–10015.
32. Yang, H.-B.; Hawkridge, A. M.; Huang, S. D.; Das, N.; Bunge, S. D.; Muddiman, D. C.; Stang, P. J. *J. Am. Chem. Soc.* **2007,** *129,* 2120–2129.
33. Yang, H.-B.; Ghosh, K.; Northrop, B. H.; Zheng, Y.-R.; Lyndon, M. M.; Muddiman, D. C.; Stang, P. J. *J. Am. Chem. Soc.* **2007,** *129,* 14187–14189.
34. Olenyuk, B.; Levin, M. D.; Whiteford, J. A.; Shield, J. E.; Stang, P. J. *J. Am. Chem. Soc.* **1999,** *121,* 10434–10435.
35. Kusukawa, T.; Fujita, M. *Angew. Chem. Int. Ed.* **1998,** *37,* 3142–3144.
36. Kusukawa, T.; Fujita, M. *J. Am. Chem. Soc.* **1999,** *121,* 1397–1398.

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

37. Ibukuro, F.; Kusukawa, T.; Fujita, M. *J. Am. Chem. Soc.* **1998,** *120,* 8561–8562.
38. Ito, H.; Kusukawa, T.; Fujita, M. *Chem. Lett.* **2000,** 598–599.
39. Fujita, M.; Oguro, D.; Miyazawa, M.; Oka, H.; Yamaguchi, K.; Ogura, K. *Nature* **1995,** *378,* 469–471.
40. Yoshizawa, M; Kusukawa, T.; Fujita, M.; Yamaguchi, K. *J. Am. Chem. Soc.* **2000,** *122,* 6311–6312.
41. Nishioka, Y.; Yamaguchi, T.; Yoshizawa, M.; Fujita, M. *J. Am. Chem. Soc.* **2007,** *129,* 7000–7001.
42. Parac, T. N.; Caulder, D. L.; Raymond, K. N. *J. Am. Chem. Soc.* **1998,** *120,* 8003–8004.
43. Beissel, T.; Powers, R. E.; Parac, T. N.; Raymond, K. N. *J. Am. Chem. Soc.* **1999,** *121,* 4200–4206.
44. Davis, A. V.; Fiedler, D.; Ziegler, M.; Terpin, A.; Raymond, K. N. *J. Am. Chem. Soc.* **2007,** *129,* 15354–15363.
45. Pluth, M. D.; Tiedemann, B. E. F.; van Halbeek, H.; Nunlist, R.; Raymond, K. N. *Inorg. Chem.* **2008,** *47,* 1411–1413.
46. Leung, D. H.; Bergman, R. G.; Raymond, K. N. *J. Am. Chem. Soc.* **2008,** *130,* 2798–2805.
47. Pluth, M. D.; Bergman, R. G.; Raymond, K. N. *J. Am. Chem. Soc.* **2008,** *130,* 6362–6366.
48. Yamauchi, Y.; Yoshizawa, M.; Fujita, M. *J. Am. Chem. Soc.* **2008,** *130,* 5832–5833.

Downloaded by OHIO STATE UNIV LIBRARIES on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch013

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

Chapter 14

Fluorescent Responsive Probes for Oligonucleotide Detection

Angel A. Martí,[1] Nicholas J. Turro[2]

[1] **Department of Chemistry, Rice University, Houston, TX 77054**
[2]**Department of Chemistry, Columbia University, New York City, New York 10027**

A wide variety of probes have been designed and synthesized for detecting oligonucleotides and polynucleotides *in vivo* and *in vitro*. Of these, molecular beacons (MBs) and binary probes (BPs) have shown particular applicability to specific problems such as mRNA tracking, single nucleotide polymorphism, and polymerase chain reaction quantization. MBs are hairpin oligonucleotide probes, containing a fluorophore and a quencher, that change their fluorescent properties upon binding to a given target. BPs, on the other hand, consist of two fluorophore-containing oligonucleotide strands that hybridize to adjacent regions of a target sequence, thus favoring energy transfer between the neighboring fluorophores. These probes have been extensively studied and modified to enhance their detection characteristics using different dye combinations, three-dye arrays, excimer-forming molecules and metal complexes. The design, applicability and advantages of these probes for the detection and tracing of oligonuclotides in different media will be discussed.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

The use and development of probes to obtain information about systems that otherwise would be impossible to study has become a major landmark of modern science and engineering (*1*). Probes to study the Earth's outer atmosphere, the solar system, as well as the depths of the ocean have revealed important information, which has expanded our understanding about these environments. We can find excellent examples of probes in nature. Our five senses allow us to probe and sense our surroundings. At the microscopic level, our immune system is constantly probing our body, in search for viruses and infections, and different organs in our body incessantly monitor the level of a variety of components in our blood stream such as sugars, lipids and electrolytes, which play a vital role in their regulation.

Artificial nanoscopic probes, based on single molecules (*2*) and/or polymers (*3*), have been synthesized for the purpose of probing different species such as ions (*4*), viruses (*5*), and DNA (*6*). These probes are usually designed so they recognize and/or interact with the species of interest, based on the complementarity of their physicochemical properties; this complementarity can arise from hydrogen bonding, complexation, hydrophobic interactions, morphological and topological considerations, etc. In the case of oligo and polynucleotide probes, the design is commonly based on the hybridization of a target DNA or RNA sequence to an artificial oligonucleotide probe made complementary to the target sequence (*6-10*) (throughout this chapter the term "oligonucleotide" will be used for short DNA or RNA strands of less than 30 bases long, whilst the term "polynucleotide" will be used for larger molecules such as full-length RNA (*6*)). These hybridization probes generally contain a reporter group (usually a fluorescent molecule), which presents some change in their physical properties when the probe binds to the target (*11*). During this chapter we will describe the design and properties of two popular hybridization probes: molecular beacons (MBs) and binary probes (BPs). This will yield a general view of how hybridization probes can be used to study and detect DNA and RNA with high specificity and sensitivity in different media.

Molecular Beacons

MBs have gained special attention since their development by Tyagi and Kramer in 1996 (*7*). A standard MB is composed of an oligonucleotide strand possessing a fluorophore and a quencher at its opposite ends (*12*). The strand is composed of a loop, which is complementary to a target DNA or RNA sequence, and a stem part composed of short self-complementary sequences at opposite sides of the strand, which keep the MB in the "closed" conformation in the absence of a target (Figure 1(a)). In this "closed" conformation, the fluorophore (F) and the quencher (Q) are close to one another and excitation of F does not result in any fluorescence, since it is quenched by energy transfer to Q (*13*). In the presence of a target, the loop and the target hybridize, disrupting the MB stem-loop conformation and spatially separating the fluorophore and the quencher, resulting in a strong fluorescence emission (Figure 1(b)).

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

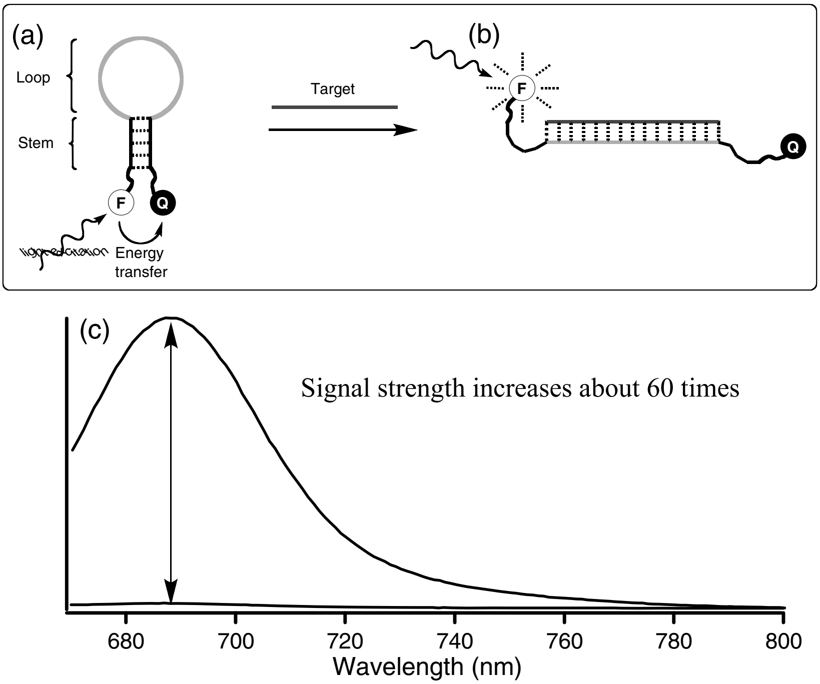


Figure 1. (a) "Classic" MB in the "closed" conformation, (b) and in the "open" conformation when hybridized with target; (c) steady-state fluorescence spectra of a MB before adding target (solid line) and after the addition of target (dashed line).

The design and synthesis of efficient MBs for the detection of DNA and RNA strains require the careful tuning of some factors:

1. *Self-complementary region (stem)* – The sequence of the MB stem determines the hairpin melting point (*7*). For example, a stem with a small number of complementary G-C base pairs would result in a low stem melting point, favoring non-specific opening of the MB and increasing the background fluorescence. On the other hand, a stem with a large number of complementary G-C base pairs may have a relatively high melting point that can slow the hybridization kinetics with the target. Care should be taken to avoid complementarity between the stem and any region in the loop, which would lead to an incorrect hairpin structure. Furthermore, the formation of the hairpin structure will also depend on the salt concentration; usually 5mM $MgCl_2$ is enough to keep the MB in the "closed" conformation when no target is present (*7*). A final general consideration concerning the stem sequence is that G residues should not be placed nearby the fluorophore, since they tend to quench its fluorescence (*14*).
2. *Anti-sense region of the MB (loop)* – In general terms the size and the sequence of the loop affects the equilibrium constant between the hairpin and the random coil conformation of the MB. It has been shown that this equilibrium constant generally decreases with the loop size and the rigidity

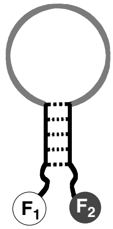
of the sequence (e.g., poly A sequences are more rigid than poly T sequences and possess smaller equilibrium constants for the same number of bases) (*15*). Ideally, some MBs 15 to 30 nucleotides long, targeting different sequences should be synthesized, and then tested to identify those that match the efficiency criteria (Signal-to-background ratio (S/B) > 10).

3. *MB reporter (fluorophore)* – A bright fluorophore is desirable since photon detection determines the S/B for any MB. The factors that favor an intense fluorescence emission are a high absorption extinction coefficient and high quantum yield (*16*), which would maximize light absorption by the MB and the fluorescence emission efficiency, respectively.
4. *"Deactivation region" of the MB (quencher)* – Several mechanisms can operate to reduce or eliminate the fluorescence by a quencher; the most common are fluorescence resonance energy transfer (FRET) and dynamic or collisional quenching (*17*). An optimum FRET quencher would have an absorption band with a high extinction coefficient overlapping the emission spectrum of the fluorophore (*18*); an optimum dynamic quencher is one that is not repelled from the fluorophore by electrostatic interactions and that can collide easily and frequently with it. The amount of dynamic quenching can be optimized by selecting the length of the linker from the fluorophore and the quencher to the MB.
5. *Complementary or anti-sense sequence of the MB (target)* – The target sequence must contain as little secondary structure as possible. This is especially important for RNA macromolecules, which are rich in secondary structure (*19*). Programs such as mFold (*20*) are used to predict the occurrence of these single-stranded regions, low in secondary structure.

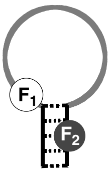
The "classical" structure of a MB has evolved into different designs with improved and extended properties (*14,21-28*). A logical step in the evolution of MBs is the use of different dye combinations to improve their detection capabilities. A brief list of examples of MB with different architectures and properties can be found in Figure 2. These designs have found applications in the detection of single nucleotide polymorphism (SNP) (*29,30*), tracing bioterrorism agents (*31*), in polymerase chain reaction (PCR) (*7,22,32*), and for cell imaging (*33,34*), among others.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

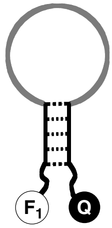
Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014



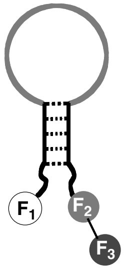
Two-dye MBs contain two fluorophores at different ends of the MB. This produces a different fluorescence output in the presence and absence of the target due to FRET. Ratiometric analysis can be used in these kinds of MBs to improve the S/B ratio (*14, 24*).



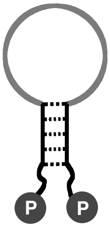
MBs with fluorescent bases do not contain fluorophores attached at both ends of the strand but fluorescent bases within the MB backbone. These fluorescent bases undergo a change in their fluorescence properties upon changes in hybridization, producing a different fluorescence signature in the absence and presence of target (*23*).



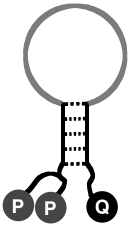
Wavelength-shifting MBs contain two fluorophores close to one another at the same end of the strand providing efficient energy transfer. Different fluorescence outputs under the same excitation source can thus be obtained by changing F_2 whilst leaving F_1 constant (*22*).



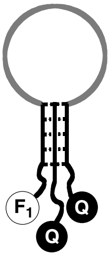
Three-dye MBs are composed of three fluorophores arranged in such a way that energy is transferred from F_1 to F_2 and subsequently to F_3 in an energy transfer cascade. Varying the distance between the fluorophores produces different spectral profiles, which allows the design of MBs with different fluorescence signatures just by varying the distance between the fluorophores (*21*).



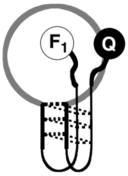
Excimer-monomer switching MBs contain pyrene groups at opposite sides of the strand. When the MB is in the absence of target, the pyrene groups are close to one another, prompting the formation of excimer emission. When the target hybridizes with the MB, the pyrene groups are apart from one another and only emission from the monomer is observed (*25*).



Multiple-pyrene labeled MBs contain two or more pyrene groups at one end of the MB and a quencher at the other end. The close distance between the pyrene groups produces excimer emission in the presence of target; in the absence of target the MB is "closed" and the excimer emission is quenched by the close proximity to the quencher (*26*).



Triplex MBs can form a triplex structure between the stem and a single strand with the right sequence. This extra strand possesses one or two quenchers, which decrease the background fluorescence in the closed MB. When the MB interacts with the complementary strand it opens and releases the extra strand containing the quencher thus producing a pronounced increase in the fluorescence observed (*27*).



Quadruplex-based MBs possess a stem sequence that allows for the formation of a quadruplex structure. This quadruplex formation is very sensitive to cation concentration allowing the modulation of the stability of the stem by changing the ionic strength of the solution (*28*).

Figure 2. Examples and description of different MBs.

Binary Probes

BPs are also hybridization probes, but in contrast to MBs, they are composed of two different oligonucleotide strands (*35,36*). A fluorophore donor is attached at one strand of the BP whilst a fluorophore acceptor is attached at the other strand. The two strands are complementary to adjacent regions of a target in such a way that upon hybridization with it, the fluorophores are brought close to one another and FRET occurs from the donor to the acceptor fluorophores (Figure 3), displaying principally acceptor fluorescence. In the absence of the target, the fluorophores are apart from one another and only fluorescence from the donor is observed (*35-42*).

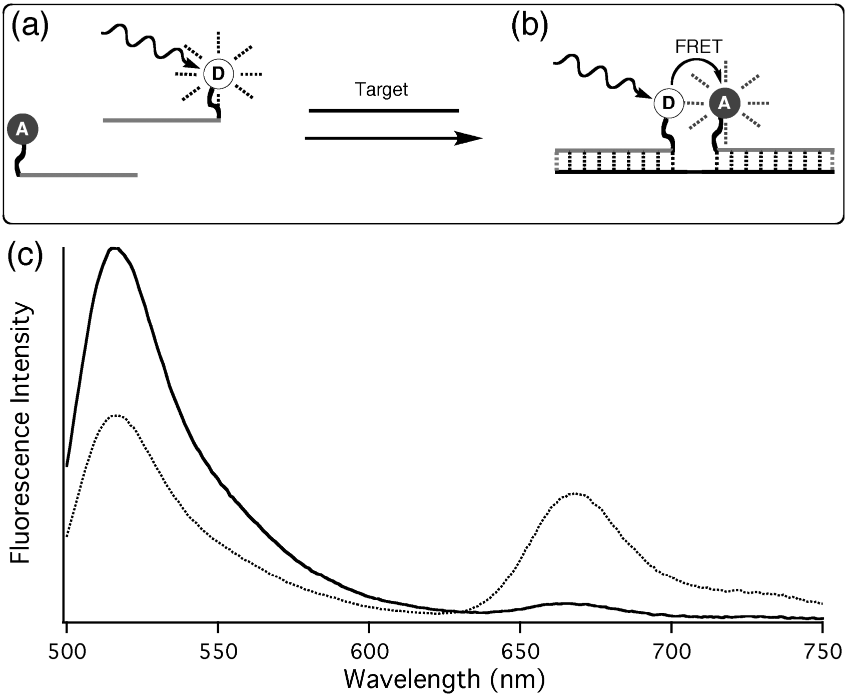


Figure 3. (a) BP before the addition of target, (b) and after the addition of target; (c) steady-state fluorescence spectra of a BP before adding target (solid line) and after the addition of target (dashed line).

As in the case of the MBs, the properties of BPs can be tuned by adjusting some parameters during their design and synthesis. Some considerations that can be taken into account in the design of BPs in order to improve their performance are:

1. *Probe sequences* – These sequences are designed such that they hybridize selectively with the target. The size of the sequence is important, i.e., very long probe sequences will increase the hybridization time, whilst relatively short ones may not produce a very stable hybrid, diminishing FRET and hence the intensity of the signal, and reducing the S/B ratio. In our experience, PNs 16 to 20 units long provide an acceptable balance between hybridization rate and stability at room temperature (*19,39*). The salt concentration is also important and usually 100 mM NaCl is appropriate for most applications. The counter cations of the salts aid in the folding of the

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

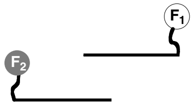
secondary structure by shielding the charge of the nucleotides phosphate groups, thus reducing the electrostatic repulsion.

2. *Fluorophores* – The fluorophores are selected to maximize the efficiency of the FRET. One of the factors to take into account to obtain efficient FRET is the overlap between the fluorescence emission of the donor and the absorption spectrum of the acceptor (the overlap integral) (*37,39*). Furthermore, since the absorption spectrum in the overlap integral is expressed in extinction coefficient units, the best acceptors tend to be those with high extinction coefficients (i.e., those that absorb light more efficiently). Another consideration is the fluorescence quantum yield of the donor; fluorophores with quantum yields near unity are the best in terms of facilitating FRET. The major factor governing FRET is the distance between the dyes (*18,39*). Theoretically, a short distance between the fluorophores should give a very efficient FRET; however if the dyes are too close to one another intermolecular collisions might quench the excited state of the donor before FRET occurs (*17*). To avoid this, BPs are usually designed with *ca.* 5 bases spacing the dyes when hybridized to target, so they are still close to allow FRET, but far enough to avoid collisions (*39*).
3. *Target sequence* - If biomolecules such as mRNA are targeted, it is important to select a region low in secondary structure, to avoid competition with the binding of the probes (*19,39*). It is also important to consider the ratio between the concentration of probe and the target. Contrary to MBs, where the target can be in excess in comparison with the amount of probe, BPs are more efficient when the probe to target ratio is close to unity (*6,39*); if the probe to target ratio changes markedly from unity, some of the probes would be distributed randomly among the target strands (<1) or free in solution (>1) and not necessarily close to one another. This would diminish FRET and consequently decrease the S/B ratio.

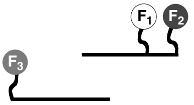
BPs have found applications in real-time monitoring of *in vivo* transcriptional mRNA synthesis (*38*), mRNA visualization in living cells (*37*), and gene translocation detection (*36*). An advantage of BPs in comparison with MBs is that they cannot present false positive detection (*6,39*). It has been observed that when MBs are injected in certain cells they present non-specific signals, i.e., they might show a positive detection signal even in the absence of the target sequence (*43,44*). This is because it is difficult to predict how MBs might interact with other components of the cell, which may cause a non-specific opening of the stem region, and detection of a false signal (*6*). Conversely, BPs do not present non-specific signals since the signal of a BP depends on both parts of the probe being in close proximity. Since the probability that this event occurs in the absence of target is very small, a positive detection signal will only arise upon specific target detection. On the other hand, some advantages of MBs over BPs are that their response times are usually faster, their detection signals are stronger, and their interactions with the target produce more stable hybrids (*6*). However, rather than compete, the properties of these probes complement each other allowing their use in a variety of applications. Figure 4 presents some examples of BPs with different structures and properties.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

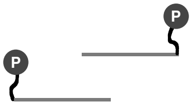
Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014



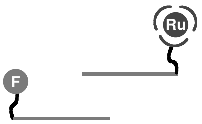
Two dye BPs are composed of two oligonucleotides, with each strand containing either a fluorophore energy donor (F_1) or acceptor (F_2). When hybridized to target, the two strands come into close contact, placing the fluorophores close to one another and promoting FRET (*37, 39*).



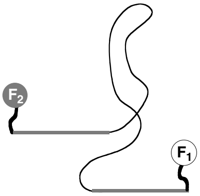
Three-dye BP architecture is formed by a three fluorophore arrangement in which energy is relayed from a fluorophore (F_1) to the acceptor (F_3) through a second fluorophore (F_2). This allows less overlap between the emission signals and a reduction in the direct excitation of the acceptor (*39*).



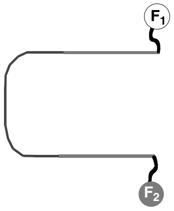
Pyrene BPs contain pyrene groups attached to each of the two oligonucleotide strands. In the absence of target the probes only present pyrene monomer emission; however, when hybridized to target, the pyrene groups are close to one another and excimer fluorescence emission is obtained (*19*).



Ru-Cy5 BPs possess a ruthenium complex (Ru) as an energy donor and a Cy5 fluorophore (F) as an energy acceptor. When the probes are free in solution only emission from the ruthenium complex is observed. When the probes are bound to the target, spin-forbidden energy transfer from the ruthenium complex triplet state to the Cy5 singlet state occurs, producing a delay in the fluorescence emission of the latter (*40*).



PEG-linked BPs use poly(ethylene glycol) to bridge together the two BP segments. The two linked segments are apart enough to prevent FRET in the absence of target. Because the two PB segments are bound together, hybridization of one of them to the target facilitates the binding of the other (*41*).



C-probes also bear the basic structure of BPs but the two probe chains are linked by a long oligonucleotide chain. This linker is large enough to preclude FRET in the absence of target; however, in the presence of target, it improves the hybridization kinetics by keeping both probe strands bound together (*42*).

Figure 4. Examples and description of different BPs.

Case Study: Oligonucleotide and Polynucleotide Detection in Highly Fluorescent Background Media

The aforementioned probes allow the detection of oligo- and polynucleotides with high selectivity and sensitivity in buffer solution or non-fluorescent media. However, as discussed above, it is also of interest to detect DNA and RNA in cellular and other biological fluids, which often contain intrinsic fluorescence (*19,40*). This fluorescence precludes the effective detection of the probes by substantially increasing the amount of background. Fluorescence lifetime measurements of cellular and biological fluids have shown short lifetimes of the order of 3 to 8 ns (*6,19,40*). The short-lived fluorescence of these media allows their discrimination with respect to the fluorescence of longer-lived probes with the use of time-resolved emission spectroscopy. This technique allows the detection of emission spectra at specific times after light excitation. Figure 5 shows the concept of time-resolved emission spectroscopy. The background of the cellular medium decays at a faster rate than the fluorescence of the probes. Taking the time-resolved emission spectrum (TRES) after the fluorescence of the background has decayed, allows us to selectively monitor the fluorescence of the probes without the interference of other fluorescent species.

Figure 6a shows the standard fluorescence spectra (steady-state spectra) for an inorganic-organic hybrid BP composed of a Ru complex as the energy donor

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

and an organic energy acceptor, an indodicarbocyanine dye (Cy5) (Same as Figure 4 Ru-Cy5 BPs) (*40*). Figure 6b shows the steady-state fluorescence spectra for the same set of probes in the presence of a strongly fluorescent cell medium background. When comparing Figures 6a and b, it is possible to distinguish that the fluorescence background in Figure 6b is dominating the spectra, which prevents the sensitive detection of the probes. Since in this probe the energy is transferred from the triplet excited state of the Ru complex to the singlet excited state of the acceptor, the transition is spin-forbidden. This has the effect of slowing the rate of energy transfer, which results in a delay in the fluorescence decay of the Cy5 energy acceptor. Because the fluorescence decay of Cy5 is now spread over a longer time scale, it is possible to detect the fluorescence of Cy5 without the interference of the fluorescent background (*40*). Figure 6c shows the TRES of Ru-Cy5 probes time-gating the detection from 59 to 77 ns. Figure 6c closely matches Figure 6a, practically eliminating all the fluorescence background observed in Figure 6b.

Time-resolved fluorescence spectroscopy is a technique allowing one to discriminate short-lived fluorescent backgrounds from probes with long-lived luminescence decays. This technique has proven efficient in the detection of RNA in cell extracts from *Aplysia californica* (*19*), platelet-derived growth factor in Dulbecco's cell-growing medium (*45*), and oligonucleotides in highly fluorescent backgrounds (*40*). Additionally, the growth of the fluorescence lifetime imaging (FLIM) field opens new opportunities for the use of these probes for *in vivo* applications (*46,47*). This technique presents an alternative for the detection of different species, especially proteins and polynucleotides, in living cells and tissues.

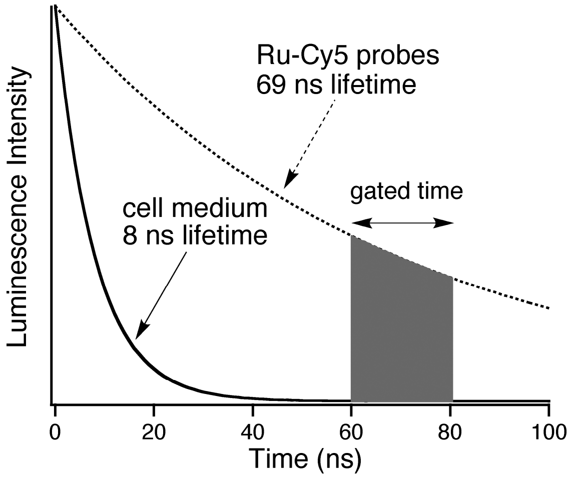
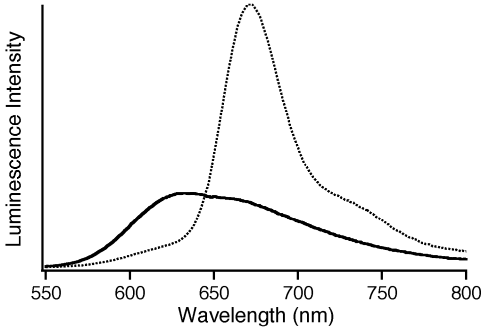


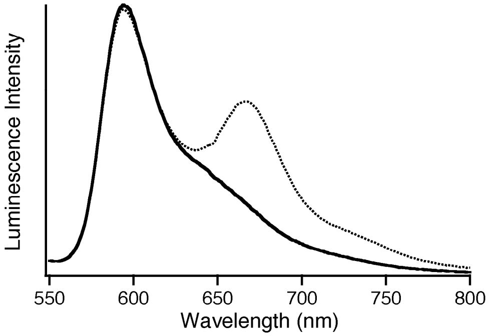
Figure 5. Fluorescence time-decay profiles for the strongly fluorescent cell medium (solid line) and Ru-Cy5 probes (dashed line). The observed fluorescence decay of Cy5 after spin-forbidden energy transfer is longer-lived than the fluorescence of the medium, allowing one to time gate the fluorescence detection to monitor the probes after the fluorescent background from the medium has decayed (for clarity, only the longer-lived components were included in the figure. For more information see reference (40)).

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

(a) Steady-state emission spectra in buffer solution



(b) Steady-state emission spectra in strongly fluorescent cell medium



(c) Time-resolved emission spectra in strongly fluorescent cell medium

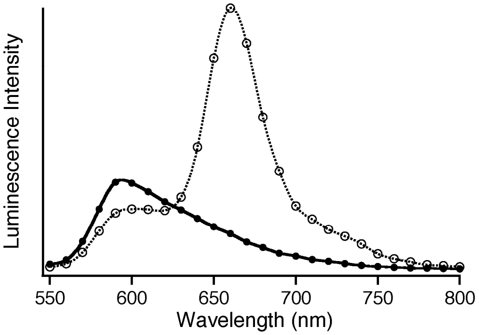


Figure 6. (a) Steady-state luminescence spectra of Ru-Cy5 probes in buffer solution, (b) in highly fluorescent cell medium, (c) and TRES from 59 to 77 ns in the absence (solid line) and presence of target (dashed line). Modified from reference (40)).

Conclusions

The detection of DNA and RNA is of great importance in areas such as molecular biology (*7,22,32,36,38*), neurobiology (*19*), medicine (*5,36,48*), and home-land security for bacterial agent detection (*31*), among others. Hybridization probes such as MBs present good sensitivity, fast response time and have been widely used in a variety of different applications which is a good indicator of their usefulness and efficiency (*6*). BPs on the other hand are more specific (they do not present non-specific detection) and are easier to synthesize (*39*). These characteristics have made MBs and BPs important tools for the detection of DNA and RNA *in vivo* and *in vitro*. Nevertheless, further improvement of these probes would be necessary to achieve routine detection of pico and femto molar concentration of target in different media. The use of more intensely fluorescent dyes such as Quantum dots (*49,50*), may be useful for the design and synthesis of more sensitive probes. Also, the use of time-resolved emission spectroscopy with long-lived dyes such as pyrene (*19,45*), luminescent transition metal complexes (*40*), and lanthanide chelates (*51*) in combination with FLIM (*46,47*), presents an important alternative for the detection of oligonucleotides and polynucleotides in strongly fluorescent media.

References

1. Mason, W. T. *Fluorescent and luminescent probes for biological activity,* 2nd ed.; Biological Techniques Series; Academic Press: San Diego, CA / London, 1999.
2. Hirano, T.; Li, W.; Abrams, L.; Krusic, P. J.; Ottaviani, M. F.; Turro, N. J. Supramolecular steric effects as the means of making reactive carbon radicals persistent. Quantitative characterization of the external surface of MFI zeolites through a persistent radical probe and a langmuir adsorption isotherm. *J. Org. Chem.* **2000,** *65,* 1319–1330.
3. Wang, D.; Carroll, G. T.; Turro, N. J.; Koberstein, J. T.; Kovác, P.; Saksena, R.; Adamo, R.; Herzenberg, L. A.; Herzenberg, L. A.; Steinman, L. Photogenerated glycan arrays identify immunogenic sugar moieties of Bacillus anthracis exosporium. *Proteomics* **2007,** *7,* 180–184.
4. Nagatoishi, S.; Nojima, T.; Juskowiak, B.; Takenaka, S. A pyrene-labeled G-quadruplex oligonucleotide as a fluorescent probe for potassium ion detection in biological applications. *Angew. Chem., Int. Ed. Engl.* **2005,** *44,* 5067–5070.
5. McKillen, J.; Hjertner, B.; Millar, A.; McNeilly, F.; Belák, S.; Adair, B.; Allan, G. Molecular beacon real-time PCR detection of swine viruses. *J. Virol. Methods* **2007,** *140,* 155–165.
6. Martí, A. A.; Jockusch, S.; Stevens, N.; Ju, J.; Turro, N. J. Fluorescent hybridization probes for sensitive and selective DNA and RNA detection. *Acc. Chem. Res.* **2007,** *40,* 402–409.
7. Tyagi, S.; Kramer, F. R. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* **1996,** *14,* 303–308.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

8. Marras, S. A. E.; Tyagi, S.; Kramer, F. R. Real-time assays with molecular beacons and other fluorescent nucleic acid hybridization probes. *Clin. Chim. Acta* **2006,** *363,* 48–60.
9. Levy, M.; Cater, S. F.; Ellington, A. D. Quantum-dot aptamer beacons for the detection of proteins. *ChemBioChem* **2005,** *6,* 2163–2166.
10. Silverman, A. P.; Kool, E. T. Quenched probes for highly specific detection of cellular RNAs. *Trends Biotechnol.* **2005,** *23,* 225–230.
11. Antony, T.; Subramaniam, V. Molecular beacons: nucleic acid hybridization and emerging applications. *J. Biomol. Struct. Dyn.* **2001,** *19,* 497–504.
12. Tan, W.; Fang, X.; Li, J.; Liu, X. Molecular beacons: a novel DNA probe for nucleic acid and protein studies. *Chem. Eur. J.* **2000,** *6,* 1107–1111.
13. Tan, W.; Wang, K.; Drake, T. J. Molecular beacons. *Curr. Opin. Chem. Biol.* **2004,** *8,* 547–553.
14. Jockusch, S.; Martí, A. A.; Turro, N. J.; Li, Z.; Li, X.; Ju, J.; Stevens, N.; Akins, D. L. Spectroscopic investigation of a FRET molecular beacon containing two fluorophores for probing DNA/RNA sequences. *Photochem. Photobiol. Sci.* **2006,** *5,* 493–498.
15. Bonnet, G.; Krichevsky, O.; Libchaber, A. Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95,* 8602–8606.
16. Lakowicz, J. R. *Principles of fluorescence spectroscopy,* 2nd ed.; Kluwer Academic/Plenum Publishers: New York, 1999.
17. Marras, S. A. E.; Kramer, F. R.; Tyagi, S. Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Res.* **2002,** *30,* e122.
18. Van Der Meer, B. W.; Coker III, G.; Chen, S. S.-Y. *Resonance energy transfer: theory and data;* VCH Publishers, Inc.: New York, 1994; pp 5–33.
19. Martí, A. A.; Li, X.; Jockusch, S.; Li, Z.; Raveendra, B.; Kalachikov, S.; Russo, J. J.; Morozova, I.; Puthanveettil, S. V.; Ju, J.; Turro, N. J. Pyrene binary probes for unambiguous detection of mRNA using time-resolved fluorescence spectroscopy. *Nucleic Acids Res.* **2006,** *34,* 3161–3168.
20. Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **1999,** *288,* 911–940.
21. Li, X.; Li, Z.; Martí, A. A.; Jockusch, S.; Stevens, N.; Akins, D. L.; Turro, N. J.; Ju, J. Combinatorial fluorescence energy transfer molecular beacon for probing nucleic acid sequences. *Photochem. Photobiol. Sci.* **2006,** *5,* 896–902.
22. Tyagi, S.; Marras, S. A. E.; Kramer, F. R. Wavelength-shifting molecular beacons. *Nat. Biotechnol.* **2000,** *18,* 1191–1196.
23. Martí, A. A.; Jockusch, S.; Li, Z.; Ju, J.; Turro, N. J. Molecular beacons with intrinsically fluorescence nucleotides. *Nucleic Acids Res.* **2006,** *34,* e50.
24. Zhang, P.; Beck, T.; Tan, W. Design of a molecular beacon DNA probe with two fluorophores. *Angew. Chem., Int. Ed. Engl.* **2001,** *40,* 402–405.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

25. Fujimoto, K.; Shimizu, H.; Inouye, M. Unambiguous detection of target DNAs by excimer-monomer switching molecular beacons. *J. Org. Chem.* **2004,** *69,* 3271–3275.
26. Conlon, P.; Yang, C. J.; Wu, Y.; Chen, Y.; Martinez, K.; Kim, Y.; Stevens, N.; Martí, A. A.; Jockusch, S.; Turro, N. J.; Tan, W. Pyrene excimer signaling molecular beacons for probing nucleic acids. *J. Am. Chem. Soc.* **2008,** *130,* 336–342.
27. Grossmann, T. N.; Röglin, L.; Seitz, O. Triplex molecular beacons as modular probes for DNA detection. *Angew. Chem., Int. Ed. Engl.* **2007,** *46,* 5223–5225.
28. Bourdoncle, A.; Torres, A. E.; Gosse, C.; Lacroix, L.; Vekhoff, P.; Saux, T. L.; Jullien, L.; Mergny, J.-L. Quadruplex-based molecular beacons as tunable DNA probes. *J. Am. Chem. Soc.* **2006,** *128,* 11094–11105.
29. Frutos, A. G.; Pal, S.; Quesada, M.; Lahiri, J. Method for detection of single-base mismatches using bimolecular beacons. *J. Am. Chem. Soc.* **2002,** *124,* 2396–2397.
30. Bonnet, G.; Tyagi, S.; Libchaber, A.; Kramer, F. R. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96,* 6171–6176.
31. Varma-Basil, M.; El-Hajj, H.; Marras, S. A. E.; Hazbón, M. H.; Mann, J. M.; Connell, N. D.; Kramer, F. R.; Alland, D. Molecular beacons for multiplex detection of four bacterial bioterrorism agents. *Clin. Chem.* **2004,** *50,* 1060–1063.
32. Tyagi, S.; Bratu, D. P.; Kramer, F. R. Multicolor molecular beacons for allele discrimination. *Nat. Biotechnol.* **1998,** *16,* 49–53.
33. Bratu, D. P.; Cha, B., -J.; Mhlanga, M. M.; Kramer, F. R.; Tyagi, S. Visualizing the distribution and transport of mRNA in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100,* 13308–13313.
34. Santangelo, P.; Nitin, N.; Bao, G. Nanostructured probes for RNA detection in living cells. *Ann. Biomed. Eng.* **2006,** *34,* 39–50.
35. Cardullo, R. A.; Agrawal, S.; Flores, C.; Zamecnik, P. C.; Wolf, D. E. Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U. S. A.* **1988,** *85,* 8790–8794.
36. Mergny, J.-L.; Boutorine, A. S.; Garestier, T.; Belloc, F.; Rougée, M.; Bulychev, N. V.; Koshkin, A. A.; Bourson, J.; Lebedev, A. V.; Valeur, B.; Thuong, N. T.; Hèléne, C. Fluorescent energy transfer as a probe for nucleic acid structures and sequences. *Nucleic Acids Res.* **1994,** *22,* 920–928.
37. Tsuji, A.; Koshimoto, H.; Sato, Y.; Hirano, M.; Sei-lida, Y.; Kondo, S.; Ishibashi, K. Direct observation of specific messenger RNA in a single living cell under a fluorescence microscope. *Biophys. J.* **2000,** *78,* 3260–3274.
38. Sei-lida, Y.; Koshimoto, H.; Kondo, S.; Tsuji, A. Real-time monitoring of in vitro transcriptional RNA synthesis using fluorescence resonance energy transfer. *Nucleic Acids Res.* **2000,** *28,* e59.
39. Martí, A. A.; Li, X.; Jockusch, S.; Stevens, N.; Li, Z.; Raveendra, B.; Kalachikov, S.; Morozova, I.; Russo, J. J.; Akins, D. L.; Ju, J.; Turro, N. J.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

Design and characterization of two-dye and three-dye binary fluorescent probes for mRNA detection. *Tetrahedron* **2007,** *63,* 3591–3600.

40. Martí, A. A.; Puckett, C. A.; Dyer, J.; Stevens, N.; Jockusch, S.; Ju, J.; Barton, J. K.; Turro, N. J. Inorganic-organic hybrid luminescent binary probe for DNA detection based on spin-forbidden resonance energy transfer. *J. Am. Chem. Soc.* **2007,** *129,* 8680–8681.
41. Yang, C. J.; Martinez, K.; Lin, H.; Tan, W. Hybrid molecular probe for nucleic acid analysis in biological samples. *J. Am. Chem. Soc.* **2006,** *128,* 9986–9987.
42. Rosmarin, D.; Pei, Z.; Blaser, M. J.; Tyagi, S. US patent, 20060040275, 2006.
43. Santangelo, P. J.; Nix, B.; Tsourkas, A.; Bao, G. Dual FRET molecular beacons for mRNA detection in living cells. *Nucleic Acids Res.* **2004,** *32,* e57.
44. Tanke, H. J.; Dirks, R. W.; Raap, T. FISH and immunocytochemistry: towards visualising single target molecules in living cells. *Curr. Opin. Biotechnol.* **2005,** *16,* 49–54.
45. Yang, C. J.; Jockusch, S.; Vincens, M.; Turro, N. J.; Tan, W. Light switching excimer probes for rapid protein monitoring in complex biological fluids. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102,* 17278–17283.
46. Carlsson, K.; Liljeborg, A. Confocal fluorescence microscopy using spectral and lifetime information to simultaneously record four fluorophores with high channel separation. *J. Microsc.* **1996,** *185,* 37–46.
47. Lakowicz, J. R.; Szmacinski, H.; Nowaczyk, K.; Berndt, K. W.; Johnson, M. Fluorescence Lifetime Imaging. *Anal. Biochem.* **1992,** *202,* 316–330.
48. Tsourkas, A.; Bao, G. Shedding light on health and disease using molecular beacons. *Brief. funct. genomics proteomics* **2003,** *1,* 372–384.
49. Kim, J. H.; Stephens, J. P.; Morikis, D.; Ozkana, M. Hybrid inorganic-organic molecular beacons. *Sensor Letters* **2004,** *2,* 85–90.
50. Kim, J. H.; Morikis, D.; Ozkana, M. Adaptation of inorganic quantum dots for stable molecular beacons. *Sens. Actuators, B* **2004,** *102,* 315–319.
51. Tsourkas, A.; Behlke, M. A.; Xu, Y.; Bao, G. Spectroscopic features of dual fluorescence/luminescence resonance energy-transfer molecular beacons. *Anal. Chem.* **2003,** *75,* 3697–3703.

Downloaded by CORNELL UNIV on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch014

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

Chapter 15

Artificial Photosynthesis

James T. Muckerman[1,2] and Etsuko Fujita[1]

[1]**Chemistry Department, Brookhaven National Laboratory,**
Upton, NY 11973
[2]**Center for Functional Nanomaterials, Brookhaven National Laboratory,**
Upton, NY 11973

Research in artificial photosynthesis seeks to replicate the natural process of photosynthesis that converts water and carbon dioxide into carbohydrates and oxygen using sunlight as the energy source. The visible-light driven splitting of water into hydrogen and oxygen and the production of methanol from CO_2 reduction are sometimes included in the definition. There are two distinct approaches to artificial photosynthesis: structural models *vs*. functional models of the natural systems. We are pursuing functional models as exemplified in recent work on hydrogenase-inspired catalysts for H_2 production. The function of the chlorophyll arrays in plants might be replaced with a suitable band-gap-narrowed semiconductor photoanode in an electrochemical cell with an attached molecular multi-electron water oxidation catalyst. The role of the NADH co-factor as the carrier and donor of two electrons and a proton might be replaced by a photogenerated hydride donor to carry out the function of Photosystem I. Here we briefly review the catalytic production of oxygen, hydrogen and methanol from water and carbon dioxide using transition-metal complexes as catalysts, and we describe progress in our recent work and that of others.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Introduction

In 2003 the worldwide energy consumption was 14 TW. By 2050 it is projected to increase at least two-fold (to ca. 30 TW) owing to economic and population growth, even after substantial decreases in energy utilization due to technological improvements in energy production, distribution, end use and conservation are considered. Solar energy is the most abundant source of energy available; the energy from sunlight that strikes the earth in only one hour is equivalent of all the energy currently consumed from various sources on the earth in a whole year! (*1*) While solar energy can been captured and utilized by biomass and photovoltaics, less than a few percent of fuels and electricity are currently provided by sunlight. Electric energy generation by photovoltaics is currently an order of magnitude more expensive than conventional energy generation using fossil fuels (*i.e.*, coal, gas, oil), but the supply of fossil fuels is undoubtedly limited. When oil and gas reserves are depleted, coal will have to be used, and the combustion of coal produces even more CO_2 than petroleum-based fuels. This will create an unprecedented level of global warming owing to a rapid increase in CO_2 emissions into the atmosphere. Ethanol production from biomass may be in principle “carbon neutral”, and has been heavily subsidized by federal/local government policies that are supporting a burgeoning industry, but it threatens food resources as production is scaled up to meet demand. Moreover, according to two recent full-cycle analyses of the global environmental cost of the production of almost all biofuels used today, they lead to even more CO_2 emission than conventional fuels (*2,3*).

Until technological breakthroughs for solar-to-fuel conversion are achieved or costs of fossil-based energy become higher, solar energy is not likely to play a large role in power generation. While photovoltaic-electrolyzer systems offer some promise, they currently require an expensive solar collector and a large amount of platinum for electrodes. Moreover, we do not have efficient enough batteries to store solar-generated electricity for night use on a global scale. These considerations have led many researchers to consider solar generation of fuels (stored in the form of chemical bonds such as hydrogen from water and methanol from CO_2) as the best and essential solution for meeting global energy needs in an environmentally responsible manner. While sunlight-driven water splitting or CO_2 reduction to methanol/methane remains a formidable problem, we need to be searching now for breakthroughs and to develop cost-effective, carbon-neutral, energy systems that produce clean and sustainable energy. Energy is an essential resource needed for solving any of the problems facing humanity, and chemists and material scientists can make a large contribution toward solving these problems during the next few decades.

Hydrogen is a clean energy source and does not emit CO_2 into the atmosphere. Hydrocarbons (fossil fuels and biomass) and water are the only two pools of hydrogen atoms on Earth that can sustain long-term, large-scale hydrogen production (*4*). The use of water requires energy (*i.e.*, for the net reaction $2\ H_2O \rightarrow 2\ H_2 + O_2$, $\Delta G° = 4 \times 1.23$ eV per electron/hole (e^-/h^+) pair = 113.4 kcal/mol of gaseous O_2 at STP), but it is also environmentally benign, does not contribute to the accumulation of greenhouse gases, and does not compete for valuable and limited hydrocarbon (or agricultural) resources.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

Furthermore, it provides CO-free H_2 for polymer electrolyte membrane (PEM) hydrogen fuel cells avoiding problematic Pt catalyst deactivation. These advantages can be fully realized only if the energy comes from a clean, renewable source such as solar energy, which is the most attractive option for the future (*1,5-7*).

The energetic requirements for water decomposition into H_2 and O_2 (eqs 1-5) or CO_2 reduction (eqs 6-12) at pH 7 *vs*. NHE (the normal hydrogen electrode) depend on the number of electrons in the redox half-reactions as shown below.

Water splitting:

$$H^+ + e^- \rightarrow H^\bullet \qquad E^o = -2.61\ V \tag{1}$$

$$2OH^\bullet + 2H^+ + 2e^- \rightarrow 2H_2O \qquad E^o = 2.31\ V \tag{2}$$

$$2H^+ + 2e^- \rightarrow H_2 \qquad E^o = -0.41\ V \tag{3}$$

$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O \qquad E^o = 0.82\ V \tag{4}$$

$$H_2O_2 + 2H^+ + 2e^- \rightarrow 2H_2O \qquad E^o = 1.35\ V \tag{5}$$

CO_2 reduction:

$$CO_2 + e^- \rightarrow CO_2^{-}\bullet \qquad E^o = -1.9\ V \tag{6}$$

$$CO_2 + H^+ + 2e^- \rightarrow HCO_2^- \qquad E^o = -0.49\ V \tag{7}$$

$$CO_2 + 2H^+ + 2e^- \rightarrow CO + H_2O \qquad E^o = -0.53\ V \tag{8}$$

$$CO_2 + 4H^+ + 4e^- \rightarrow C + 2H_2O \qquad E^o = -0.20\ V \tag{9}$$

$$CO_2 + 4H^+ + 4e^- \rightarrow HCHO + H_2O \qquad E^o = -0.48\ V \tag{10}$$

$$CO_2 + 6H^+ + 6e^- \rightarrow CH_3OH + H_2O \qquad E^o = -0.38\ V \tag{11}$$

$$CO_2 + 8H^+ + 8e^- \rightarrow CH_4 + 2H_2O \qquad E^o = -0.24\ V \tag{12}$$

While one-electron reduction of a proton and CO_2 (eqs 1 and 6, respectively) and oxidation of water (eq 2) take place at very negative and positive potentials, respectively, the coupled multi-electron and multi-proton reactions occur at relatively modest potentials (eqs 3-5, 7-12).

Hydrogen from water can be driven with solar light either by using a photovoltaic device that generates electricity for conventional electrolysis or by combining the water oxidation and reduction processes in a so-called photoelectrolysis device (Figure 1) or removing the wires from the second system in an artificial photosystem as described in Figure 1 in reference 1. In the second and third devices we must incorporate several key elements, including: (a) an apparatus for light absorption and charge separation that uses the energy of an absorbed photon to generate a one-electron oxidant and reductant (shown as hole and electron, respectively), prevents their recombination, and directs

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

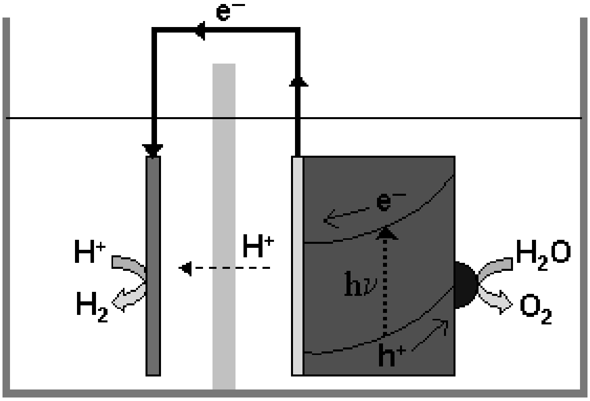


Figure 1. Photoelectrolysis cell with a BGNSC photoanode (right) and a catalyst at the anode surface. A small bias potential may be required in some cases to produce H_2 at a cathode electrode with a catalyst.

them toward spatially-separated catalysts; (b) a reduction catalyst that makes molecular hydrogen in a two-electron redox process; and (c) an oxidation catalyst that promotes the transfer of four electrons from two water molecules to the incoming holes, thereby producing an oxygen molecule. To drive the water splitting, the amount of photon energy stored in the form of redox potentials must be at least 1.23 eV per charge separation event; in practice, it is substantially larger. Extensive research efforts over the past 30 years have revealed that these stringent requirements for photochemical water splitting are extremely difficult to meet. The challenging problems are: (1) the efficient collection of photons from the solar spectrum; (2) the efficient charge separation with visible light in materials that are stable to photocorrosion; (3) the four-electron water oxidation catalysis; (4) the production of hydrogen by non-noble metal catalysts; and (5) the efficient coupling of (1)-(4) without a substantially large energy loss or overpotential.

In the first 30 years following the initial discovery of photoelectrolysis using *n*-type TiO_2 (*8*), the search for improved semiconductors focused mainly on binary oxides and simple ternary oxide systems (*9*). The common problem with these first-generation oxides is their wide band gaps, in excess of 3 eV, that makes them unusable with visible light (with a 3 eV band gap the maximum theoretical energy conversion efficiency approaches only 3%) (*10,11*). This 3 eV band-gap ceiling is much larger than the thermodynamic 1.23 V potential required for water splitting reactions, and this roadblock occurs because most of the initial strategies used to reduce band gaps only served to lower the conduction band energy. In the past several years, it has been demonstrated that solar water splitting can be carried out with band-gap-narrowed semiconductors (BGNSCs) such as N- (*12-21*), C- (*22-25*), or B-doped (*26-28*) TiO_2, or other oxynitrides (*29-44*) and oxysulfides (*45-47*) that have been prepared for use as photocatalysts. A number of outstanding claims concerning the photocatalytic properties of these new materials have been made that have not yet been independently confirmed, especially in the case of photochemical water splitting

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

(*16,22,48-52*). The most promising of these materials to date is $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ (*36-44,53*), which has been reported to have a quantum efficiency of about 50% for O_2 production when illuminated by visible light in the presence of a sacrificial electron acceptor (*38*). However, this efficiency drops to 2.5% in the wavelength range of 420 to 440 nm when the sacrificial reagent is replaced by a proton reduction catalyst. Thus the greatest current challenge is to design improved high-efficiency BGNSCs for solar-driven water splitting by understanding their structural and electronic properties (*53*) and their light-harvesting and charge-separation ability, and to search for those with the conduction-band position very close to the two-electron proton reduction potential for efficient proton reduction catalysis.

A TiO_2 photoanode does not require a water oxidation catalyst because the very high reduction potential of its holes can support even the energetically wasteful one-electron oxidation of water to the OH radical ($E^o = 2.31$ eV at pH 7). This will no longer be the case for BGNSCs suitable for efficient solar water splitting applications. Undoubtedly, when the valence band (VB) edge is optimized for solar water splitting, a catalyst that can utilize holes with this reduction potential for four-electron water oxidation as in Figure 1 will be required. On the other hand, in order to utilize electrons in the conduction band (CB) of semiconductors, platinum group metals have often been used as catalysts for hydrogen generation. However, if hydrogen is to become viable as a fuel, low-cost, robust catalysts will be needed. Recent advances have demonstrated that functional bio-inspired hydrogenase-model complexes with pendent bases in the outer coordination sphere of the metal increase rates and decrease the overpotentials associated with electrocatalytic hydrogen production and oxidation.

Natural photosystems (Figure 2) in green plants convert CO_2 to carbohydrates and O_2 using absorbed photons as energy and water as a reducing agent. The light energy absorbed by chlorophylls results in charge separation. In Photosystem I , using available protons, photogenerated electrons are converted to chemical energy in the form of adenosine triphosphate (ATP) and reduced nicotinamide adenine dinucleotide phosphate (NADPH) in which a reduced hydrogen equivalent (*i.e.*, hydride) is stored by the reduction of NAD^+. The ATP and NADPH are used to reduce CO_2 in the light-independent, complicated reactions of the Calvin cycle through net hydride-ion transfer reactions. While TiO_2 and TiO_2-grafted zeolites have been used to produce CO together with a trace amount of CH_3OH and CH_4 under UV irradiation, products in homogeneous photo- and electrochemical reduction of CO_2 reported so far have been limited to CO and/or HCOOH. Photogenerated holes are utilized to oxidize water at the oxygen-evolving center (OEC) of Photosystem II. The OEC contains a Mn_3CaO_4 cluster linked to a fourth Mn by a μ-oxo bridge (*54*) and is capable of turning over about 10^3 O_2 molecules/s. (*55*). So far no man-made catalysts (including metal oxides, and manganese and ruthenium complexes) can generate O_2 in such an efficient way.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

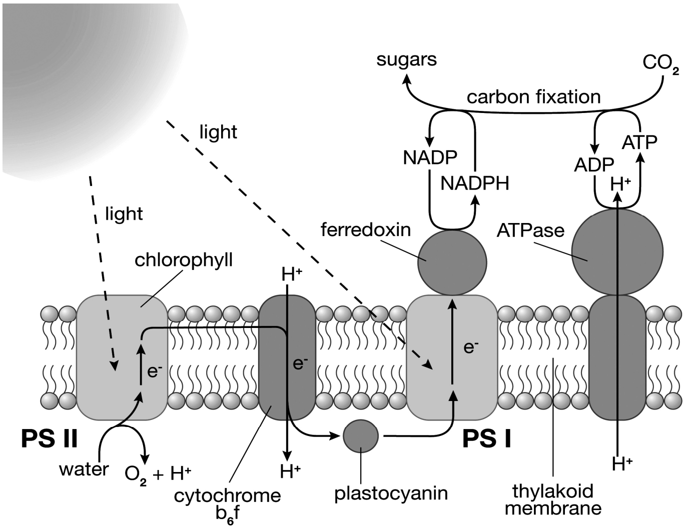


Figure 2. Schematic diagram of natural photosynthesis.

In this review we discuss our view of materials for use as a BGNSC photoanode and the dark catalysis for capturing photoinduced electrons/holes from such a photoanode (or any other potential photon absorbers to produce electricity), and for converting abundant materials such as water and CO_2 to fuels. There are many articles that describe photochemical water splitting by semiconductor materials together with simple metal or metal oxide catalysts such as RuO_2, Rh, NiO. Here we focus on molecular catalysts and discuss only the following several subjects along with our personal views about what we need to pursue in the upcoming decade as chemists who are interested in developing efficient, robust and non-noble-metal catalysts for water splitting and CO_2 reduction, search for sustainable pathways for energy generation and contribute to the most important, yet challenging problem facing the world in this century.

- Visible light absorption and charge separation using band-gap-narrowed semiconductor (BGNSC) materials that are stable in an oxidizing, aqueous solution environment and suitable for use as a photoanode.
- Water oxidation by ruthenium molecular catalysts instead of the manganese O_2-forming clusters in Photosystem II. (We will omit the manganese O_2-forming molecular catalysts since most of these catalysts require O atoms from sacrificial oxidants to produce O_2.)
- H_2 production and oxidation by bio-inspired hydrogenase-like catalysts containing inexpensive metals.
- Photogeneration of renewable hydride donors, carriers of a proton and two electrons, to replace the $NADP^+$/NADPH coenzyme for the reduction of CO_2 and related species.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

Visible Light Absorption and Charge Separation Using Band-Gap-Narrowed Semiconductors

The solar energy required to drive the water splitting process must be absorbed by a chromophore to produce a charge-separated state, and those separated charges must be coupled into the water oxidation and proton reduction half-reactions. Natural photosynthesis accomplishes this using chlorophyll molecules and a sequence of only slightly exothermic electron-transfer reactions. An alternative approach for artificial systems is to use semiconductors that absorb visible light to create electron/hole pairs. When the photoanode in a photoelectrochemical cell is made of such a semiconductor material, band bending to achieve a match of chemical potentials at the aqueous interface and the electrode on opposite sides of the semiconductor drives the hole to migrate to the aqueous interface of the semiconductor and the electron to the electrode side, as illustrated in Figure 1.

In practice, things are not as simple as described above. There are three important criteria that a semiconductor material must meet in order to be useful as a photoanode in a photoelectrochemical cell: (1) it must absorb visible light, *i.e.*, it must have a band gap that is smaller than 3.00 eV but at least as large as 1.65 eV (the thermodynamic threshold of 1.23 eV at pH 7 plus some overpotenital); (2) the band edges of the semiconductor, *i.e.*, the top of the valence band and the bottom of the conduction band, must straddle the electrochemical potentials for the four-electron oxidation of water to molecular oxygen and the two-electron reduction of protons to molecular hydrogen; and (3) the semiconductor material must be stable with respect to corrosion in an oxidizing aqueous electrolyte solution environment. Transition metal oxides such as TiO_2, and $SrTiO_3$ straddle the potentials for water oxidation and proton reduction and are stable in aqueous solution, but have wide band gaps that allow them to absorb only UV light that comprises *ca.* 3% of the solar spectrum. Indeed, researchers some time ago successfully used such semiconductors to split water, but with a very low efficiency (*8,9,56-65*). Other semiconductors such as CdS (which is employed in light meters for cameras because it absorbs so well in the visible region of the solar spectrum) and CdSe have almost ideal electrical and optical properties for water splitting applications, but they are extremely unstable in aqueous solution. In fact, there are no known semiconductors that have been demonstrated to completely meet all three criteria listed above as a single-component photo-absorber or anode.

The search for a semiconductor material that is appropriate for solar water splitting has been continuing for over twenty years, and has focused primarily on doping wide band-gap metal oxide semiconductors with cations and/or anions in an effort to reduce their band gaps. This goal is made more difficult to achieve by the fact that the conduction band edge of most of the candidate metal oxides is barely adequate for the two-electron reduction of protons so that reducing its band gap by lowering the conduction band does not produce a useful material. In recent years there have been promising reports of selectively raising the edge of the valence band of certain metal oxides though anion doping or the formation of solid solutions of different semiconductor materials (*12-52,66*).

Anion-Doped Titania

Using X-ray photoelectron spectroscopy (XPS), near edge X-ray absorption fine structure (NEXAFS), and X-ray diffraction (XRD) we have characterized N-doped TiO_2 prepared by several methods: an ion implantation on the TiO_2(110) surface (*67*); a film prepared by molecular beam epitaxial growth (*68,69*); N-doping of titania by NH_3 at above 600 °C (*70*); and the introduction of nitrogen into TiO_2 nanoparticles prepared by a sol-gel method in solution containing $N(C_2H_5)_3$ below 90 °C. While N-doped TiO_2 prepared by ion implantation shows N substitution for O in the lattice together with O vacancies, the yellow materials prepared in solution with $N(C_2H_5)_3$ do not show N substitution (*i.e.*, no TiN-like N 1*s* peak at 396 eV in XPS). In general, our experimental and theoretical studies show a low stability of N inside the titania lattice and a difficulty in generating the large concentrations of the dopant that are necessary to produce useful changes in the band gap of the oxide. In several cases, we could not confirm results reported in the literature using the same sample preparation recipes. New studies of the recently described oxynitride alloy GaN/ZnO provide, in contrast, promising evidence of controllable and substantial stable doping for both high quality materials and reduced bandgap (*38*).

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

GaN/ZnO Solid Solutions

Because of the low N-substitution level and the lack of photocatalytic activity toward water oxidation in our N-doped TiO_2, we changed our focus toward an oxynitride species with more promising photocatalytic activity, GaN/ZnO (*66*). Recently, a solid solution of GaN and ZnO, $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$, has been reported to be a stable photocatalyst that is capable of water splitting under visible light (*33-44*). $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ has a single-phase hexagonal wurtzite-type crystal structure similar to pure GaN and ZnO. The lattice constants, band gap, and the overall activity of this photocatalyst for water splitting have been shown to depend strongly on the zinc concentration, x (*71,72*) However, the x range studied experimentally is quite limited (generally $x < 0.25$) and the nature of the observed band-gap narrowing is not well understood. We therefore initiated coordinated experimental and theoretical studies of $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$.

We have carried out the first systematic theoretical study on the variation of the structural and electronic properties of the $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solution with the Zn (or equivalently, O) mole fraction, x, by using density functional theory (DFT) to obtain the composition dependence of the physical and optical properties of this material and a detailed understanding of the band-gap-narrowing mechanism (*53*). The two major challenges of this study were the arrangement of the constituent atoms in the solid solution at different concentration, x, and the fact that standard DFT (GGA and LDA) methods do a poor job of predicting the band gap of metal oxides, and do a particularly poor job for ZnO owing to the strong O(2*p*)-Zn(3*d*) coupling. To address the first of these challenges, we have adopted the special quasirandom structures (SQS) approach of Zunger *et al.* (*73*) in which special small-unit-cell periodic structures are constructed to closely mimic the most important structural features of the corresponding random solid solution. To address the deficiency of standard DFT in predicting band gaps of metal oxides that complicate the calculations on GaN/ZnO solid solutions from the outset, we have adopted the GGA+U approach (*74-76*), where U is an on-site Coulomb interaction correction for the Ga and Zn 3d orbitals. We plot the calculated band gaps E_g using the most stable configurations as a function of x for both 16-atom $(Ga_{8-n}Zn_n)(N_{8-n}O_n)$ and 32-atom $(Ga_{16-n}Zn_n)(N_{16-n}O_n)$ supercells in Figure 3.

Owing to the use of the empirical U correction, our calculated band gaps are qualitatively correct but are still underestimated compared to the experimental values. Although there is scatter in the calculated band gaps owing to the finite size of the supercells employed, over the entire composition range we see that the band gap, E_g, exhibits a downward bowing as a function of Zn (O) concentration for both supercells, which suggests that there is a minimum band gap for the solid solution at some intermediate Zn concentration. For $A_{1-x}B_x$ alloys, the band gap $E_g(x)$ can be described by (*77,78*).

$$E_{\mathrm{g}}(x) = (1-x)\cdot E_g(\mathrm{A}) + x \cdot E_{\mathrm{g}}(\mathrm{B}) - b \cdot x \cdot (1-x)$$

where b is an optical bowing parameter. For our pseudobinary $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solution, this bowing parameter can be estimated *via* a least-squares fit to the combined data obtained for the two types of supercells. Using experimentally measured end member (pure GaN and ZnO) published band gap values (*71*) and our theoretically estimated bowing parameter, the $E_g(x)$ behavior of the $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solution should follow the downward-bowed dashed curve shown in Figure 3. We see that again our results are in qualitative agreement with experimental findings for experimentally accessed mole fraction of $x < 0.25$. The minimum band gap is estimated to be about 2.29 eV at $x = 0.525$, suggesting that efforts to synthesize materials with $x > 0.25$ may be rewarded with increased photocatalytic activity for visible light.

In order to realize the enhanced properties predicted by our DFT calculations, we have carried out preliminary studies of the reaction mechanisms by which $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ is formed with the goal of synthesizing more highly doped (larger x) materials. We prepared a series of solid solutions by the reaction of Ga_2O_3 and ZnO with NH_3 and followed the transformation with *in-situ* and *ex-situ* methods, including XPS, NEXAFS, and XRD. *In-situ* XRD patterns (Figure 4) during the synthesis of $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solutions from a 1:1 Ga_2O_3/ZnO mixture with NH_3 indicate an intermediate phase of $ZnGa_2O_4$ with a spinel structure (diffraction peaks at 17° and 20°) (*79*). In separate experiments we prepared $ZnGa_2O_4$ by heating a mixture of β-Ga_2O_3 and ZnO to 800–850 °C. The $ZnGa_2O_4$ proved to be an excellent precursor for the synthesis of $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solutions with good crystallinity. Our results of XANES and XPS studies indicate that the ZnO trapped in the GaN lattice is electronically perturbed with respect to pure ZnO. The content of ZnO in $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ decreases when the compound is exposed to NH_3 at 850–900 °C. Thus, we have found a methodology to control the amount of ZnO inside GaN by controlling the exposure of the precursor materials to NH_3. We

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

plan to study the structure of $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solutions in detail using a combination of EXAFS, high-energy XRD, neutron diffraction, and X-ray and neutron PDF techniques.

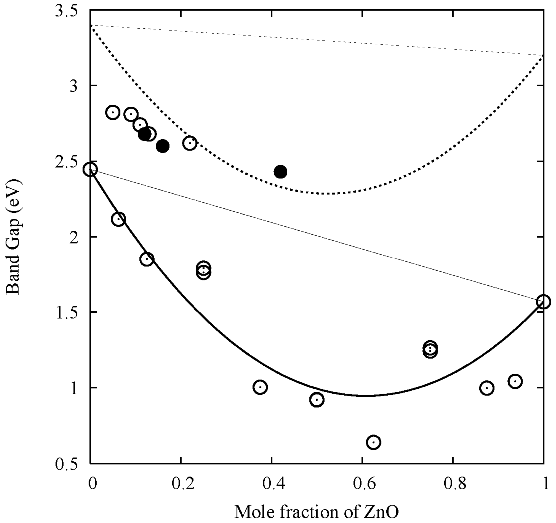


Figure 3. Variation of band gap as a function of Zn (O) concentration, x. Lower points and curve (solid): calculated BGs and smoothed $E_g(x)$ curve using the estimated bowing parameter, b. Upper points and curve (dashed): experimental data for $(Ga_{1-x}Zn_x)(N_{1-x}O_x)$ solid solution (38,71) and predicted experimental $E_g(x)$ behavior using the estimated b and the limiting GaN and ZnO band gaps (71).

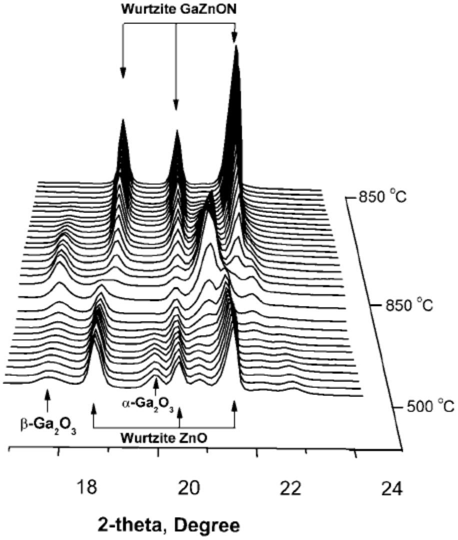


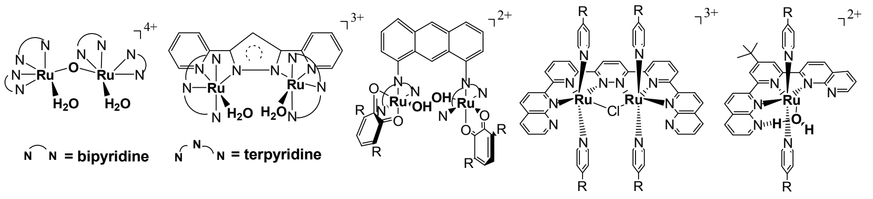
Figure 4. In-situ XRD patterns for the nitridation of a 1:1 Ga_2O_3/ZnO mixture by 5% NH_3 in He. The peaks at ca. 17° and 20° arise from the intermediate spinel phase.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

Water Oxidation Catalysis

Examples of Water Oxidation Catalysts

Cyanobacteria, algae, and green plants reduce CO_2 to carbohydrates and oxidize water to O_2 using visible light, thereby converting solar energy into chemical energy. While Nature can supply the inspiration for man-made systems to achieve similar functionality, it is very difficult to design systems to carry out the photoinduced oxidation and reduction reactions in such continuously regenerating and elegant ways (*80,81*). The platinum group metals (Ru, Ir, Pt, Os) have been studied as water oxidation catalysts, and with some optimization colloidal $IrO_x{\cdot}nH_2O$ (*82*) has exhibited an O_2 evolution rate of 40 molecules s^{-1} per surface Ir atom using a sacrificial oxidant such as $Ru(bpy)_3^{3+}$ (bpy = 2,2′-bipyridine). By contrast, as mentioned in the Introduction, the Mn_3CaO_4 cluster in the oxygen-evolving complex (OEC) in Photosystem II is capable of turning over at a remarkable rate (about 1000 O_2 molecules s^{-1}) (*55*). While many binuclear transition-metal catalysts of the type $L(H_2O)M—O—M(OH_2)L$ or $L(H_2O)M(BL)M(OH_2)L$ (where L and BL are non-bridging and bridging organic ligands, respectively) have been prepared and investigated as potential water oxidation catalysts (*83-111*), even the best of them cannot evolve O_2 faster than one O_2 molecule s^{-1} per catalyst molecule (*81,83,91,93,112*). Some of the ruthenium catalysts of this type are shown in Figure 5. The complexes containing the Ru—O—Ru motif (**1**) together with bipyridine derivatives are reported to have catalytic activity for water oxidation. Interestingly, a few mononuclear $Ru—OH_2$ complexes such as$Ru(binapy)(4\text{-}R\text{-}py)_2(H_2O)]^{2+}$ (**5**, binapy = 2,6-bis(2′-naphthyridyl)-4-*t*-butylpyridine, R = Me, NMe_2, etc.) have also been reported to produce O_2 upon the addition of an acidic Ce^{IV} solution (*111*), but the mechanisms are not clear and their redox, spectroscopic, and reactivity properties need to be further investigated.



1 **2** **3** **4** **5**

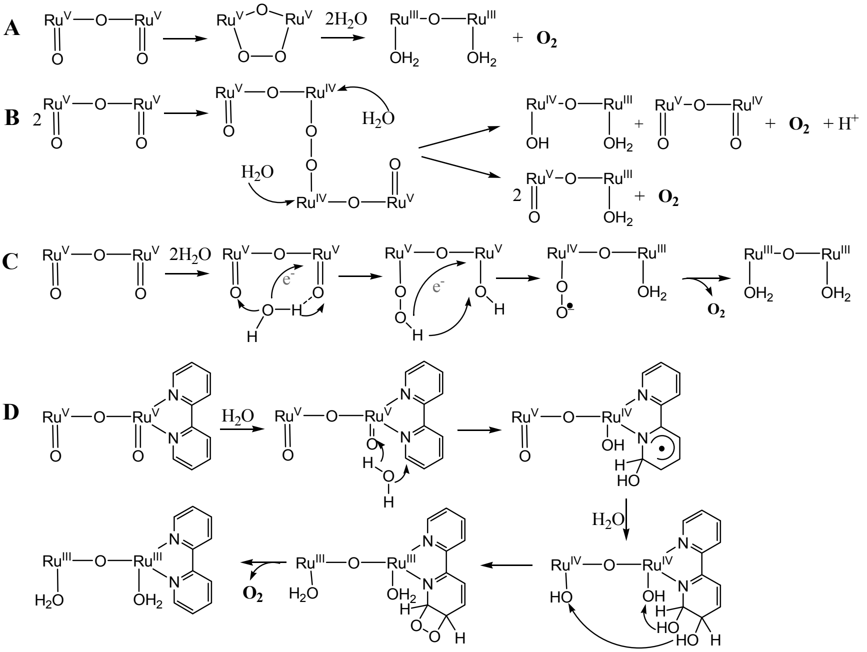
Figure 5. Structures of several Ru water oxidation catalysts

The Most Thoroughly Investigated "Blue Dimer"

Impressive progress has been achieved in unraveling the molecular mechanism of the catalytic action of the so-called blue dimer catalyst (*86-99,113,114*) $[cis,cis\text{-}(bpy)_2(H_2O)Ru—O—Ru(OH_2)(bpy)_2]^{4+}$ (**1**) by combining structural and kinetic studies. However, the detailed mechanism of the final stage of O—O bond formation and O_2 evolution remain unclear even after 30

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

years have passed since the initial discovery of water oxidation by this complex. Furthermore, there are some inconsistent results on rates of individual proton-coupled electron-transfer steps (or disproportionation steps), and mechanistic questions in this extensively investigated system still exist (Scheme 1).



Scheme 1. Various mechanisms proposed for O_2 evolution from the blue dimer.

It is generally believed that [*cis,cis*-$(bpy)_2(H_2O)Ru$—O—$Ru(OH_2)(bpy)_2]^{4+}$ (**1**, Figure 5) here denoted as $(H_2O)Ru^{III}$—O—$Ru^{III}(H_2O)$, is oxidized by four holes and the removal of four protons to form O=Ru^{V}—O—Ru^{V}=O, which is believed to be the active form of the catalyst that produces O_2 (mechanism A in Scheme 1). Hurst *et al.* carried out a series of labeling experiments with excess Ce(IV) in 0.5 M HSO_3CF_3 under conditions that produce O=Ru^{V}—O—Ru^{V}=O, monitoring the isotopic composition of the O_2 produced. Using $(H_2{}^{18}O)Ru^{III}$—O—$Ru^{III}(H_2{}^{18}O)$ and $H_2{}^{16}O$ solvent, they found comparable formation of $^{16}O^{18}O$ and $^{16}O^{16}O$ and negligible $^{18}O^{18}O$ during the first catalytic turnover, independent of temperature. The mechanistic pathway for the appearance of $^{16}O^{18}O$ has been proposed (by the groups of Hurst and Meyer (*91,93,113,114*) and examined using DFT calculations by Baik *et al.* (*95*)) to involve the attachment of one molecule of the water solvent to one of the Ru^{V}=O moieties (mechanism C in Scheme 1). Although this mechanism can explain the appearance of $^{16}O^{18}O$, it does not explain the $^{16}O^{16}O$ formation. An isotopic ligand exchange with the solvent ($H_2{}^{16}O$) is unlikely because Hurst *et al.* did not observe a temperature dependence of the isotopic composition of the O_2 evolution. Therefore, Hurst *et al.* recently suggested mechanism D (Scheme 1) involving covalent hydration of O=Ru^{V}—O—Ru^{V}=O to produce a transient (O=Ru^{V}—O—Ru^{IV}—OH)(bpyOH$^{\bullet}$)

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

ligand radical intermediate, which is attributed to the NIR absorption at ~ 740 nm and ESR signal at g = 2.005 (*115*). Further characterization of this species is necessary.

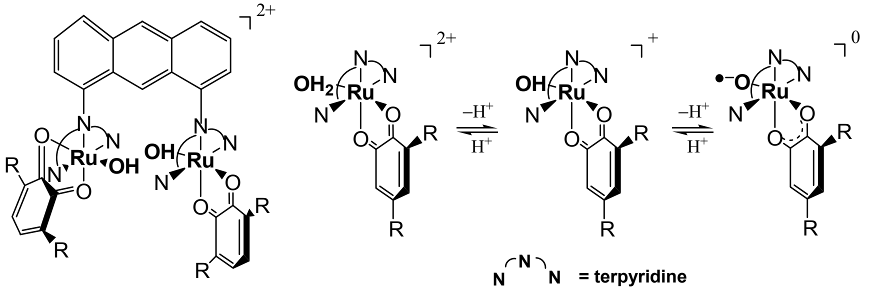
Ruthenium Complexes with Non-Innocent Quinone Ligands

Tanaka and co-workers have meanwhile reported water oxidation catalytic activity of a novel dinuclear Ru complex, $[Ru_2(OH)_2(3,6\text{-}Bu_2Q)_2(btpyan)](SbF_6)_2$ ($3,6\text{-}Bu_2Q$ = 3,6-di-*tert*-butyl-1,2-benzoquinone, btpyan = 1,8-bis(2,2′:6′,2″-terpyrid-4′-yl)anthracene) (*83,84*), that, unlike all other catalytic dinuclear Ru species, contains redox-active quinone ligands. Two $[Ru(OH)(3,6\text{-}Bu_2Q)]^+$ units are complexed by an anthracene-bridged bis-terpyridine ligand (btpyan) with a geometry in which the two OH groups on the Ru centers are in close proximity for making an O—O bond (**6**, Figure 6). The quinone ligand is generally regarded as being an electrochemically non-innocent ligand that can take on three different redox states classified as quinone (Q), semiquinone (SQ), and catecholate (Cat). An O_2 evolution turnover number of ~ 30,000 per catalyst molecule in 40 hrs has been reported for electrolysis carried out in water (pH = 4.0) with this dinuclear Ru-quinone complex with an anthracene bridge complex deposited on the surface of an indium-tin-oxide (ITO) electrode (*83,116*). This indicates that this catalyst is more rapid and more stable than other molecular ruthenium catalysts that have been studied (*81,91,93,112*).

It is useful in understanding O—O bond formation to begin with a discussion of the chemistry of the mononuclear species that comprise the dinuclear ruthenium species. Tanaka and coworkers (*117*) prepared the mononuclear aqua Ru complex $Ru(OH_2)(3,5\text{-}Bu_2Q)(tpy)^{2+}$ ($3,5\text{-}Bu_2Q$ = 3,5-di-*tert*-butyl-1,2-benzoquinone, tpy = 2,2′:6′,2″-terpyridine, see **7**, Figure 6) . with a redox-active ligand, and investigated its acid-base and redox properties. When they prepared an anthracene-bridged dinuclear complex as a water oxidation catalyst, they used $3,6\text{-}Bu_2Q$ = 3,6-di-*tert*-butyl-1,2-benzoquinone in order to avoid any complications arising from possible isomers. However, in this discussion we will henceforth simply denote both cases as Q unless otherwise specified.

The assignment of oxidation states of Ru (and Os) complexes containing a quinone ligand is difficult owing to multiple metal oxidation states, three different redox states of quinone, and various spin multiplicities, and as a result they have been ambiguous even using data such as X-ray structures and UV-vis, X-ray photoelectron, and EPR spectra (*118-126*). Even a detailed analysis of DFT-calculated structures, spin densities, and **g**-tensor anisotropies led to ambiguous (*i.e.*, intermediate) assignments.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015



6 **7** **8** **9**

Figure 6. Structure of the Tanaka catalyst and acid-base reaction of the monomer unit.

The complex **7,** $[Ru^{II}(OH_2)(Q)(tpy)](ClO_4)_2$ (*i.e.*, $[Ru(OH_2)(Q)]^{2+}$), was previously described as $[Ru^{III}(OH_2)(SQ)(tpy)](ClO_4)_2$ based on XPS data (*117*), but we now believe this complex should be assigned as $[Ru^{II}(OH_2)(Q)(tpy)](ClO_4)_2$ (*i.e.*, $[Ru(OH_2)(Q)]^{2+}$) based on UV-vis spectroscopy, the X-ray single-crystal structure, and DFT/ab initio calculations (*116*). A strong metal-to-quinone charge-transfer band at 600 nm in acidic aqueous solution shifts to 576 nm upon the addition of base. Further addition of base results in the appearance of another band at 870 nm with a concomitant decrease in the intensity of the 576 nm band. The pK_as of $[Ru^{II}(OH_2)(Q)]^{2+}$ and $[Ru^{II}(OH)(Q)]^+$ are determined to be 5.5 and 10.7, respectively, by the spectroscopic method.

Tanaka and coworkers (*117*) proposed that the ruthenium oxyl radical complex $[Ru^{II}(O^{\bullet-})(SQ)]^0$ was prepared by the double deprotonation of the aqua ligand of "$[Ru^{III}(OH_2)(SQ)]^{2+}$" through "$[Ru^{III}(OH)(SQ)]^{+}$" as shown in Scheme 2 *with revised assignments* (*116*) by the first two sets of arrows.

$$[Ru^{II}(OH_2)(Q)]^{2+} \underset{+H^+}{\overset{-H^+}{\rightleftharpoons}} [Ru^{II}(OH)(Q)]^{+} \underset{+H^+}{\overset{-H^+}{\rightleftharpoons}} [Ru^{II}(O^{\bullet -})(SQ)] \xrightarrow{H^{\bullet}} [Ru^{II}(OH)(SQ)]$$

7 **8** **9** **10**

Scheme 2. Base Titration of $Ru^{II}(OH_2)(Q)(tpy)^{2+}$

However $[Ru^{II}(O^{\bullet -})(SQ)]^0$ is a very reactive oxyl radical, it is likely to abstract an H atom from CF_3CH_2OH (added to dissolve the species) to form $[Ru^{II}(OH)(SQ)]^0$. In fact, electrochemical measurements strongly suggest that in the dimeric species the oxyl radicals $[Ru^{II}(O^{\bullet -})(SQ)]^0$ may *react with each other* to form an O—O bond if the formation conditions are appropriate.

The most remarkable thing about Scheme 2 is that the oxidation state of the Ru center remains Ru(II) while the quinone ligand acts as an electron reservoir to become a semiquinone. This is quite different from what is observed for the bpy analogue investigated by Takeuchi, *et al.* (*127*), as shown in the base titration of $Ru^{II}(OH_2)(bpy)(tpy)^{2+}$, **11**, in Scheme 3. The pK_a of $[Ru^{II}(OH_2)(bpy)]^{2+}$ is about 10, which is much larger than that of

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

$[\mathrm{Ru^{II}(OH_2)(Q)}]^{2+}$ (pK_a 5.5). While a proton-coupled oxidation to form $[\mathrm{Ru^{IV}(O)(bpy)}]^{2+}$, **13**, can be accomplished by applying potentials smaller than 0.8 V for pH between 3 and 10, the formation of $[\mathrm{Ru^{IV}(O)(Q)}]^{2+}$ from $[\mathrm{Ru^{III}(O)(Q)}]^{2+}$ is not coupled to a proton transfer and requires a more positive potential (1.1 V) (*116*). This could lead to different behavior of these two species as water oxidation catalysts.

$$\underset{\mathbf{11}}{[\mathrm{Ru^{II}(OH_2)(bpy)}]^{2+}} \xrightleftharpoons[+\mathrm{H^+},\ +\mathrm{e^-}]{-\mathrm{H^+},\ -\mathrm{e^-}} \underset{\mathbf{12}}{[\mathrm{Ru^{III}(OH)(bpy)}]^{2+}} \xrightleftharpoons[+\mathrm{H^+},\ +\mathrm{e^-}]{-\mathrm{H^+},\ -\mathrm{e^-}} \underset{\mathbf{13}}{[\mathrm{Ru^{IV}(O)(bpy)}]^{2+}}$$

Scheme 3. Base Titration of $\mathrm{Ru^{II}(OH_2)(bpy)(tpy)^{2+}}$

As we described for water oxidation by the blue dimer in the beginning of this section, it is generally considered that $\mathrm{(H_2O)Ru^{III}—O—Ru^{III}(OH_2)}$ is oxidized by four holes and removal of four protons to form $\mathrm{O{=}Ru^{V}—O—Ru^{V}{=}O}$, which produces O_2 by one or more of the several possible mechanisms shown in Scheme 1. Furthermore $\mathrm{O{=}Ru^{V}—O—Ru^{IV}OH}$ is also believed to oxidize water.

Despite the strikingly novel features of the Tanaka catalyst with quinone ligands in comparison to the properties of its Ru analogues with the bpy ligand replacing quinone, the difficulty of preparation and the complicated electronic structures of the Tanaka catalyst have until recently prevented investigations of it outside the Tanaka group. Unlike the blue dimer (**1**), $[\mathrm{Ru_2(OH)_2(Q)_2(btpyan)}](\mathrm{SbF_6})_2$ is not soluble in water. The O—O distance between the Ru—OH units in $[\mathrm{Ru_2(OH)_2(Q)_2(btpyan)}]^{2+}$ was strategically designed using an appropriate bridge to be in close proximity for making an O—O bond from the two oxyl radicals that resulted from the deprotonation of the Ru—OH units by base titration. The electronic spectrum of this complex in MeOH showed a strong band at 576 nm, which is indicative of a metal-to-quinone charge-transfer band. Gradual addition of one equivalent of *t*-BuOK to the solution resulted in the gradual appearance of a new band at 850 nm and the disappearance of the band at 576 nm, which is consistent with the equation below and the possibility of O—O bond formation.

$$[\mathrm{Ru(OH)(Q)Ru(OH)(Q)}]^{2+} \xrightleftharpoons[2\mathrm{H^+}]{-2\mathrm{H^+}} [\mathrm{Ru(O)(SQ)Ru(O)(SQ)}]^{0}$$

Computational Studies for Elucidation of the Mechanism

Although there still exist some mechanistic questions and some inconsistencies in theoretical results from different levels of theory, we summarize here our new theoretical and experimental studies aimed toward understanding the complicated electronic and geometric structures of the Tanaka catalyst, and we discuss new directions for kinetic and mechanistic

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

investigations with the dinuclear analogues to address key issues and challenges in water oxidation. The proposed mechanism based on our calculations (*116*) (Figure 7) is *consistent with experimental observations***.** The formation of the experimentally isolated singlet species, $[Ru_2(OH)_2(Q)_2(btpyan)]^{2+}$, which is an asymmetrical hydrogen-bonded species with an O—O distance of 2.63 Å, occurs most likely through the displacement of O_2 from an intermediate such as $[(Ru)_2(O_2)(Q)_2(btpyan)]^{4+}$ by OH^- ions (or H_2O molecules followed by deprotonation) to restart a catalytic cycle. At some point in the removal of the two protons from the experimentally isolated species (which is rendered ambiguous by spin contamination problems in the UB3LYP calculations and failure of UMP2 calculations of open-shell species to converge) an inter-system crossing *(ISC) takes the complex into a series of species with higher spin multiplicity* for the rest of the catalytic cycle. These are predominantly triplet species, but the very first deprotonation is likely to lead to a quintet intermediate. It is most likely that *all protons are removed to form the near-catecholate species,* $[Ru_2(O_2^-)(Q^{-1.5})_2(btpyan)]^0$*, before the complex is oxidized.* Surprisingly, unlike other reported Ru-dinuclear water oxidation catalysts including the well-characterized blue dimer, the *predominant formal oxidation state of the two Ru atoms remains unchanged at +2* during the entire catalytic cycle according to our gas-phase calculations. Instead of an effective charge neutralization by a sequence of proton removals and one-electron oxidations (*i.e.*, $[(bpy)_2(OH_2)Ru^{III}—O—Ru^{III}(OH_2)(bpy)_2]^{4+} \rightarrow [(bpy)_2(OH)Ru^{IV}—O—Ru^{III}(OH_2)(bpy)_2]^{4+}$, etc.),(*87,91,93,113,114*) the removal of protons from $(Ru^{II})_2(OH)_2(Q)_2(btpyan)]^{2+}$ takes place with a change in total charge to form $[Ru_2(O_2^-)(Q^{-1.5})_2(btpyan)]^0$. Here the *active redox couples are SQ/Q and Cat/SQ* (consistent with the analogous complex with bpy ligands being inactive under similar conditions). While the two-electron oxidation of this species occurs at a moderate 0.4 V, the troublesome step is the next oxidation. Since it is not a proton-coupled electron-transfer reaction, the application of a relatively high potential (1.34 V *vs.* Ag/AgCl) is needed to form a species such as $[(Ru^{II})_2(O_2^{\bullet -})(Q)_2(btpyan^{\bullet +})]^{4+}$ (Figure 7), but that species is probably not the unstable anthracene cation radical. It is clear that $Ru(bpy)_3^{3+}$, a frequently used sacrificial oxidant, cannot efficiently oxidize the $[(Ru^{II})_2(O_2^{\bullet -})(Q^{-0.5})_2(btpyan)]^{2+}$ species. However, this catalyst immobilized on an ITO electrode is very stable and produces a large amount of O_2 without decomposition. This may be due to the charge stabilization/distribution in the catalyst through a site-to-site interaction between metal, quinone and the coordinated water moiety.

We are continuing both our experimental and theoretical studies to better characterize the catalytic reaction pathway of the Tanaka catalyst (*e.g.*, by employing higher-level *ab initio* calculations to resolve the ambiguity in the spin multiplicity and redox assignments) and to explore modifications of it (*e.g.*, a non-noble metal replacing the ruthenium centers and/or iminobenzoquinone replacing the quinone ligands) for improved performance in stability and overpotential.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

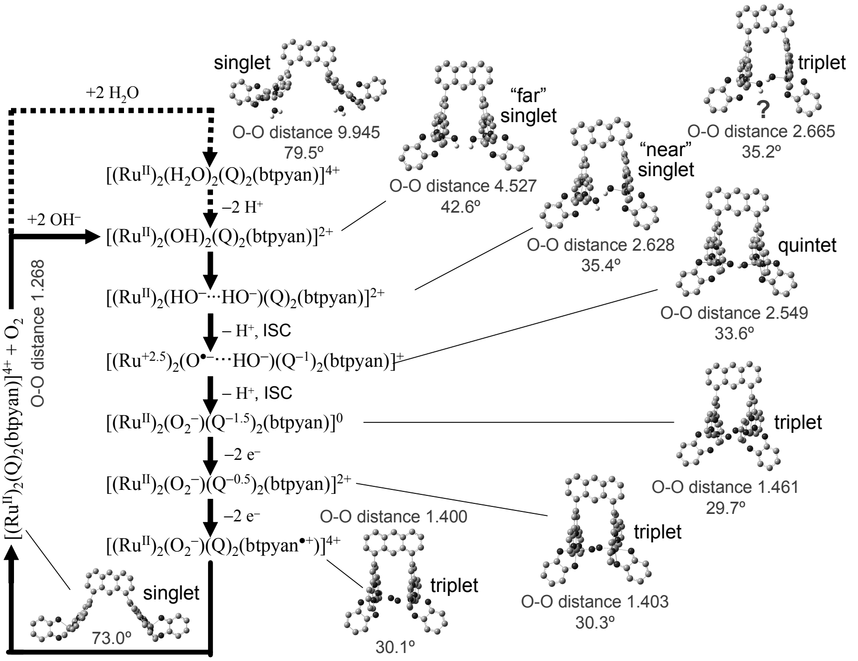

+2 H_2O
singlet
O-O distance 9.945
79.5°
$[(Ru^{II})_2(H_2O)_2(Q)_2(btpyan)]^{4+}$
$-2\ H^+$
+2 OH^-
$[(Ru^{II})_2(OH)_2(Q)_2(btpyan)]^{2+}$
"far" singlet
O-O distance 4.527
42.6°
$[(Ru^{II})_2(HO^-\cdots HO^-)(Q)_2(btpyan)]^{2+}$
"near" singlet
O-O distance 2.628
35.4°
triplet
?
O-O distance 2.665
35.2°
$-\ H^+$, ISC
$[(Ru^{+2.5})_2(O^{\bullet-}\cdots HO^-)(Q^{-1})_2(btpyan)]^+$
quintet
O-O distance 2.549
33.6°
$-\ H^+$, ISC
$[(Ru^{II})_2(O_2^-)(Q^{-1.5})_2(btpyan)]^0$
triplet
O-O distance 1.461
29.7°
$-2\ e^-$
$[(Ru^{II})_2(O_2^-)(Q^{-0.5})_2(btpyan)]^{2+}$
triplet
O-O distance 1.403
30.3°
$-2\ e^-$
O-O distance 1.400
$[(Ru^{II})_2(O_2^-)(Q)_2(btpyan^{\bullet+})]^{4+}$
triplet
30.1°
$[(Ru^{II})_2(Q)_2(btpyan)]^{4+} + O_2$
O-O distance 1.268
singlet
73.0°

Figure 7. Proposed mechanism of water oxidation by the Tanaka catalyst based on B3LYP/LANL2DZ calculations.

Hydrogen Generation and Oxidation Catalysis by Non-Noble Metal Complexes

Producing a clean fuel from water is an attractive way to eliminate the CO poisoning problem in PEM (polymer electrolyte membrane) hydrogen fuel cells or to reduce the use of fossil fuels for both economical and environmental reasons. Since Pt is expensive owing to its limited supply, here we need to investigate other materials incorporating inexpensive metals as the cathode or as catalysts for proton reduction to form H_2.

The Fe-only hydrogenase enzyme (Figure 8) generates H_2 at pH = 7 close to the thermodynamic potential of −0.41 V *vs*. NHE (or −0.65 V *vs*. SCE). In fact, this enzyme catalyzes the redox equilibrium $2H^+ + 2e^- \leftrightarrow H_2$ in both directions. The existence of nickel and iron in hydrogenase enzymes suggests that fast and efficient non-noble metal catalysts for hydrogen production can be developed. While catalysis for H_2 production/oxidation with low overpotentials has remained elusive, the structural information based on the recent X-ray diffraction studies (**14** in Figure 8) on the enzyme has led to the preparation and characterization of many dinuclear species with CO and CN ligands that mimic the structure and function of the active center (*128-139*). However, structural biomimetic model complexes only catalyze H_2 evolution at significantly more negative potentials (*ca.* −1.2 *vs.* SCE) (*128-139*). DuBois *et al.* prepared a series of *functional* models of hydrogenases in which a mononuclear Ni center and a pendent N atom act as a hydride bonding center and a proton shuttle, respectively, to lower the barrier for H_2 production/oxidation (*140-143*).

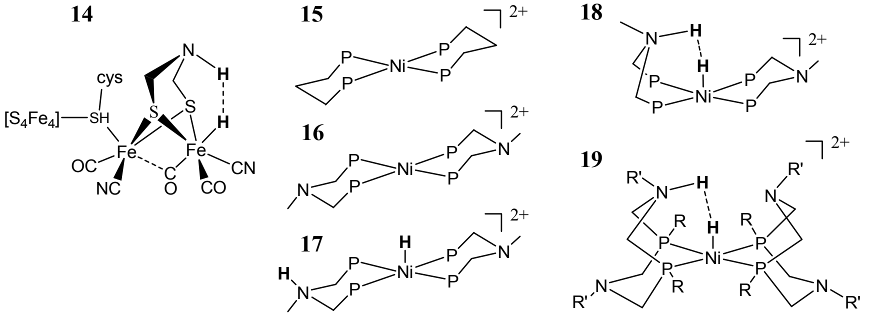


*Figure 8. Structure of Fe-only hydrogenase **14**; Ni(diphosphine)$_2$ complexes without and with a pendant base **15** and **16**, respectively; the resulting complex **17** upon addition of H_2 to **16**; a more geometrically favorable complex **18** than **17**; and an electrocatalyst **19** for H_2 production in acidic acetonitrile solutions. (R and R′ are Ph, Cy, Bz, etc.)*

They reported (*141*) that: (1) the complex **15** without any pendent base slowly oxidizes H_2 with an overpotential of ~0.73 V (using −0.14 V *vs.* Fc/Fc^+ (ferrocene/ferrocenium) as the thermodynamic potential for H^+/H_2 in CH_3CN); (2) introduction of pendent N bases into the backbone of the diphosphine ligands (complex **16**) causes a rapid heterolytic addition of H_2 to form **17**; and (3) the electrocatalytic ability of **16** is improved at low overpotentials (less than 0.08 V and a turnover frequency of H_2 oxidation between 0.01 and 0.5 s^{-1} under 1 atm of H_2) presumably through an intermediate such as **18** as predicted by our DFT calculations.

A complete catalytic cycle for hydrogen production with a simplified version of the DuBois $Ni^{II}(PNP)_2^{2+}$ catalyst (**16**, in Figure 8) as calculated in the gas phase using the B3LYP hybrid DFT method and the 6-31G** all-electron basis is shown in Figure 9 (*141*). Here all alkyl groups on the PNP ligands are replaced with hydrogen atoms. The energy of each species (in kcal/mol) relative to the bare catalyst (**I**) is given below its calculated structure. The large negative change in energy in the protonation and reduction steps reflects the quantum chemical convention of assigning zero energy to the free proton and the free electron in the gas phase. Following an initial protonation and reduction, the catalyst complex is seen to undergo a thermal oxidative addition of the proton on the protonated pendent N atom to the reduced metal center *via* the transition state shown in structure **IV**. In this structure, the ring of the active PNP ligand is seen to have flipped into a boat conformation from its original chair conformation, and remains in the boat conformation for the rest of the catalytic cycle, as suggested by structure **18** in Figure 8. This change in conformation allows a closer approach of the pendent base to the metal center, but the energy required to achieve it contributes to the activation energy of this step of the mechanism.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

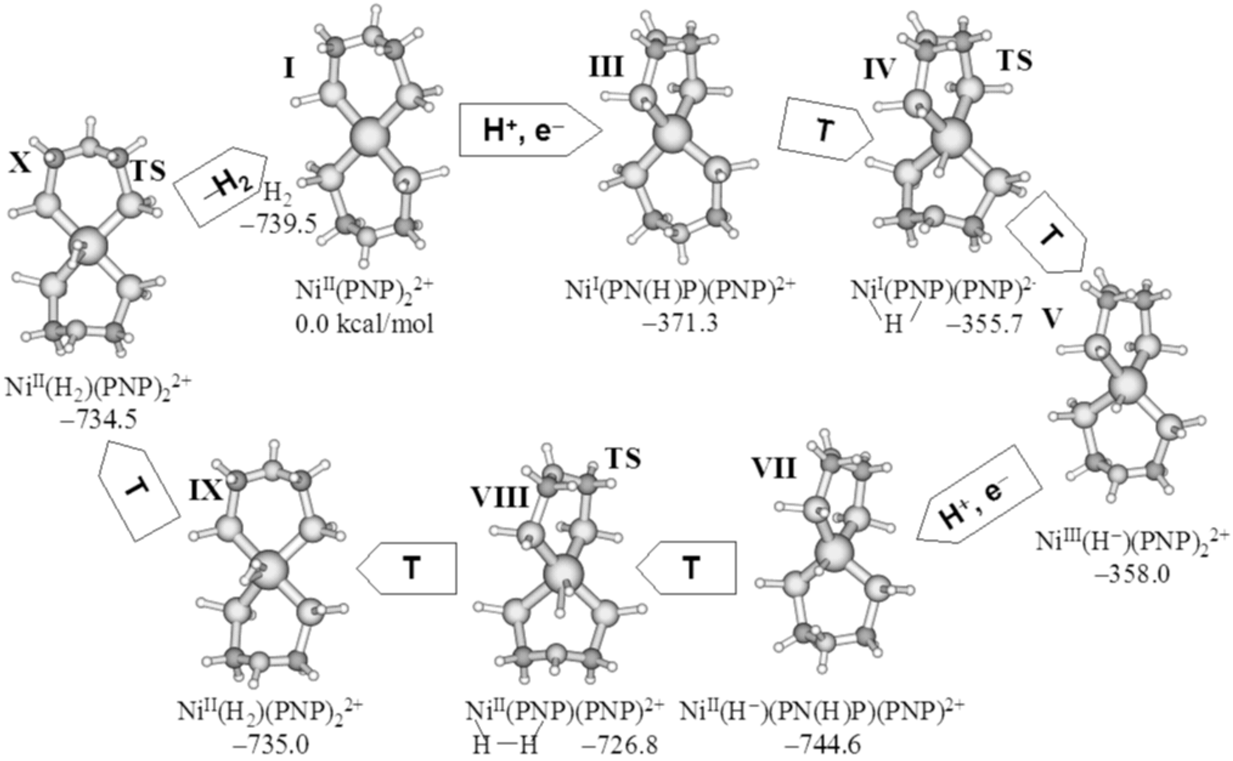

X
TS
$-H_2$
H_2
−739.5
I
$Ni^{II}(PNP)_2^{2+}$
0.0 kcal/mol
H^+, e^-
III
$Ni^{I}(PN(H)P)(PNP)^{2+}$
−371.3
T
IV
TS
$Ni^{I}(PNP)(PNP)^{2\cdot}$
H
−355.7
T
V
$Ni^{III}(H^-)(PNP)_2^{2+}$
−358.0
H^+, e^-
VII
$Ni^{II}(H^-)(PN(H)P)(PNP)^{2+}$
−744.6
T
VIII
TS
$Ni^{II}(PNP)(PNP)^{2+}$
H — H
−726.8
T
IX
$Ni^{II}(H_2)(PNP)_2^{2+}$
−735.0
T
$Ni^{II}(H_2)(PNP)_2^{2+}$
−734.5

*Figure 9. Catalytic cycle for H_2 production with DuBois PNP model catalyst based on gas-phase B3LYP/6-31G** calculations (see text). Arrows with"T" indicate thermal reactions.*

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

To avoid this unfavorable contribution to the activation energy, DuBois *et al.* further optimized the position of the N base, locking the PNP rings into the boat conformation by the use of cyclic ligands with bulky substituents (L) (*143*). The structure of the $[NiL_2(CH_3CN)]^{2+}$ complex indicated that the Ni atom is five- coordinate with a distorted trigonal bipyramidal geometry. The two N atoms in the PNP rings are in boat conformations and folded toward the Ni atom similar to the case of the heterolytic H_2 adduct shown as structure **19**. They reported a turnover frequency of 130 mol of H_2 per mole of catalyst per second in 0.1 M triflic acid at –0.94 V *vs*. Fc/Fc$^+$ applied potential (*i.e., ca.* –0.55 V *vs.* SCE).

Recent work by Peters *et al.* (*144,145*) provides a counterargument to the necessity for an outer sphere proton relay. Their work in acetonitrile solution reexamined the electrocatalytic properties of cobaloxime complexes including $Co^{II}(dmgBF_2)_2(H_2O)_2$ ($dmgBF_2$ = (difluoroboryl)dimethylglyoxime) that Espenson *et al.* reported to catalyze the reduction of protons to H_2 by Cr^{2+} in acidic aqueous solution (*146*). In CH_3CN solution, solvent molecules readily replace the axial water molecules. The cobaloxime complexes with R = CH_3 and C_6H_5 have Co(II/I) potentials at –0.55 and –0.28 V *vs.* SCE, respectively, in CH_3CN and they remarkably catalyze H_2 evolution with the small overpotentials of 40 and 50 mV, respectively, in the presence of moderately strong acids. Consistent with these small overpotentials, Peters *et al.* have shown experimentally that $Co^{II}(dmgBF_2)_2(H_2O)_2$ can catalyze both the reduction of protons and the oxidation of H_2, however, the complexes with more positive Co(II/I) potentials exhibited lower catalytic activity for H_2 production.

It should be noted that while H_2 production and oxidation with very small overpotentials have been achieved using inexpensive Ni and Co complexes, these reactions were investigated in CH_3CN, not in water. Can these results be translated to the case of water as a solvent toward future applications in electrolyzers and PEM fuel cells? Detailed kinetic and mechanistic studies in aqueous media are needed to fully understand the factors controlling efficient and low overpotential H_2 production.

Photogeneration of Renewable Hydride Donor

Photochemical conversion of CO_2 to fuels or useful chemicals using renewable solar energy is an attractive solution both to the world's need for fuels and the necessity to reduce the emission of greenhouse gases. In past work we and others have investigated photocatalytic or electrocatalytic CO_2 reduction using cobalt, nickel, ruthenium and rhenium complexes as catalysts to produce CO and/or formate (*147-182*).

We found that while the photoproduction of the reduced catalysts is fast (< 10 μs), the rate constants for CO_2 binding to the reduced metal centers depend on the catalysts and solvent, and range over ten orders of magnitude from 10^{-2} to 10^8 M^{-1} s^{-1}. The detailed mechanism of CO formation from the CO_2 adducts remains unclear in most cases, and the production of CO is very slow with a turnover frequency typically equal to or less than 10 h^{-1} even though the optimized quantum yield for CO formation is as high as 0.59 under low-

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

intensity light (*171*). Despite substantial effort, to the best of our knowledge no one has succeeded in producing methanol from CO_2, protons and electrons or H_2 using homogeneous systems. Products in the homogeneous photochemical reduction of CO_2 reported so far have been limited to CO and/or HCOOH (*156*). In the homogeneous electrochemical reduction of CO_2, a few additional C_2 species have been reported as minor products (*183*).

However, the stoichiometric chemical conversion of CO_2 to methanol or methane using $NaBH_4$ has been reported, *e.g.*, $Ru—CO_2 \rightarrow [Ru—COOH]^+ \rightarrow [Ru—CO]^{2+} \rightarrow [Ru—CHO]^+ \rightarrow [Ru—CH_2OH]^+$ (Ru = $Ru(bpy)_2(CO)$) (*184)* and $[Re—CO]^+ \rightarrow Re—CHO \rightarrow Re—CH_2OH \rightarrow Re—CH_3$ (Re = Re(Cp)(NO)(CO)) (*185*). It is therefore of considerable interest to explore whether $NaBH_4$ can be replaced by a renewable hydride. Can the photocatalytic reduction of CO_2 be carried out using metal complexes that can photogenerate hydride donors?

A number of kinetic studies of hydride transfer and electron and hydrogen-atom transfer with NADH and its model complexes have been carried out because it acts as a reservoir/source of two electrons and a proton. The NAD^+/NADH redox couple has been drawing considerable attention because chemical reactions inspired by natural systems that avoid the production of high-energy intermediates or unwanted byproducts promise to be key factors in realizing sustainable energy sources. Large numbers of photochemical reactions mediated by NADH model compounds have been extensively studied. These include the reduction of alkyl halides, olefins, ketones, and photoinduced electron transfer as well as thermal reactions using NADH analogs. However, most of the reactions mediated with NADH models have been limited to stoichiometric reactions. Tanaka's group recently prepared a polypyridylruthenium complex with an NAD^+/NADH model ligand, $[Ru(bpy)_2(pbn)]^{2+}$ ($[\mathbf{20}]^{2+}$, bpy = 2,2′-bipyridine, pbn = 2-(2-pyridyl)-benzo[*b*]-1,5-naphthyridine), which acts as a catalyst in the electrochemical reduction of acetone to 2-propanol similar to the enzymatic NAD^+/NADH (*186*). While the detailed mechanism remains unclear, this was the first example that an NADH model complex such as $[Ru(bpy)_2(pbnHH)]^{2+}$ ($[(\mathbf{20})HH]^{2+}$) presumably functions as a catalytic hydride donor for the formation of 2-propanol.

We have investigated (*187,188*) the photochemical and electrochemical properties of $Ru(bpy)_2(pbn)^{2+}$ ($[\mathbf{20}]^{2+}$), including the acid-base properties of its ground state, excited state and one-electron-reduced species. The metal-to-ligand charge-transfer (MLCT) excited state $*[\mathbf{20}]^{2+}$ lives 140 ns in CH_3CN and can be reductively quenched by an amine to produce the one-electron-reduced species. When CH_3CN solutions containing $[\mathbf{20}]^{2+}$ and triethylamine are irradiated with visible light (< 590 nm), $[(\mathbf{20})HH]^{2+}$ is produced cleanly (Φ_{355nm} = 0.21). The molecular structure of $[(\mathbf{20})HH](PF_6)_2{\cdot}2CH_3CN$, prepared by the reaction with $Na_2S_2O_4$, has been determined by x-ray diffraction, and B3LYP (and TD-B3LYP) calculations of the structure and electronic spectrum of it and the other species shown in Figure 10 agree with the experimental characterization. Reduction of $[\mathbf{20}]^{2+}$ by $CO_2^{\bullet-}$ generated by pulse radiolysis produces pH-dependent transient absorption spectra that point toward formation of the one-electron-reduced species, $[\mathbf{20}]^+$, and its protonated analogue, $[H(\mathbf{20})]^{2+}$. The pK_a of $[H(\mathbf{20})]^{2+}$ is 11. A bimolecular decay of the transient species $[H(\mathbf{20})]^{2+}$ was observed as a major reaction pathway, suggesting that the species reacts *via* disproportionation to form $[\mathbf{20}]^{2+}$ and $[(\mathbf{20})HH]^{2+}$ (Figure 10). While the $[(\mathbf{20})HH]^{2+}$ species is not a sufficiently strong hydride donor to reduce CO_2, we are currently exploring the photogeneration of other related hydride donors that have greater "hydricity" with the goal of applying them to CO_2 reduction. In summary, this type of system will allow us not only to investigate excited-state, proton-coupled-electron transfer reactions with a number of reductive quenchers that can also provide protons, but also to explore potential catalytic hydride (or hydrogen atom) transfer reactions from the hydrogenated product toward CO_2 reduction.

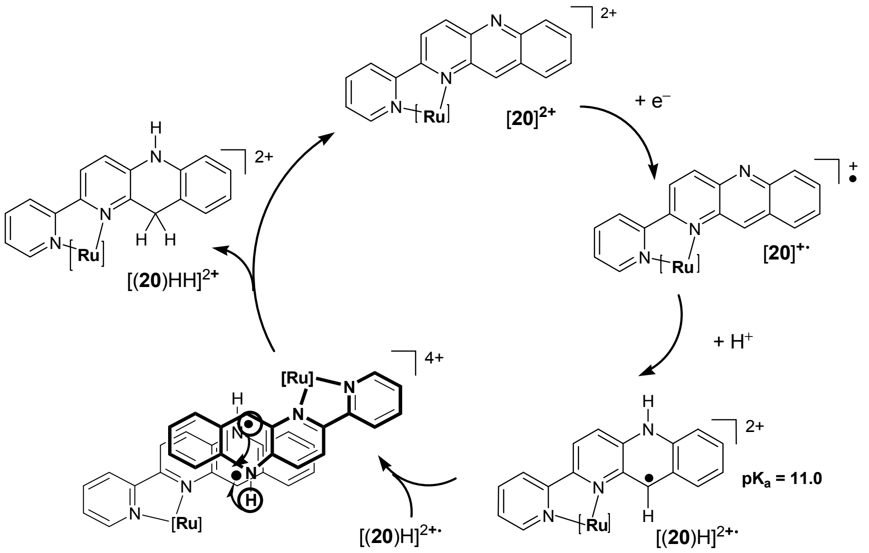


Figure 10. The mechanism of formation of final $[(\mathbf{20})HH]^{2+}$ product from the singly reduced species. Water may be a source of H^+ at high pH.

Conclusions

We have briefly reviewed our view of materials for use as a band-gap-narrowed semiconductor (BGNSC) photoanode and the dark catalysts for capturing the photoinduced electrons/holes for catalyzing the production of hydrogen, oxygen, and possibly methanol from water and carbon dioxide using transition-metal complexes in homogeneous solution, and we have described progress in our own recent work together with that of others in these areas. This includes the elucidation of hydrogen production with hydrogenase-inspired models and cobalt macrocycles, the visible light absorption and charge separation in BGNSC materials, the water oxidation mechanism of the Tanaka catalyst and the well-known blue dimer, and work toward CO_2 reduction *via* ionic hydrogenation routes using a visible-light-generated NADH-like hydride donor. Success in these areas is difficult because the feedstock molecules are extremely stable – *i.e.*, they are the products of hydrocarbon combustion,

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

respiration, etc. – and it only makes sense to use renewable sources of energy to drive their thermodynamically uphill reactions. Solar energy is clearly the most promising source of renewable energy for this purpose, but fundamental advances in coupling the solar photophysics of light absorption and charge separation to the catalytic chemical reactions that produce fuels are required to achieve this goal.

Acknowledgement

The authors gratefully acknowledge fruitful discussions with Professor Koji Tanaka (Institute for Molecular Science), Dr. Tohru Wada (Institute for Molecular Science), Dr. Dan DuBois (Pacific Northwest National Laboratory), Dr. Dmitry E. Polyansky (BNL), and Dr. Jose Rodriguez (BNL). The work at Brookhaven National Laboratory is funded under contract DE-AC02-98CH10886 with the U.S. Department of Energy and supported by its Division of Chemical Sciences, Geosciences, & Biosciences, Office of Basic Energy Sciences. We also thank the U.S. DOE for funding under the BES Hydrogen Fuel Initiatives and Solar Energy Utilization initiative.

References

1. Lewis, N. S.; Nocera, D. G. *Proc. Natl. Acad. Sci. USA* **2006,** *103,* 15729–15735.
2. Fargione, J.; Hill, J.; Tilman, D.; Polasky, S.; Hawthotne, P. *Science* **2008,** *319,* 1235–1238.
3. Searchinger, T.; Heimlich, R.; Houghton, R. A.; Dong, F.; Elobeid, A.; Fabiosa, J.; Tokgoz, S.; Hayes, D.; Yu, T.-H. *Science* **2008,** *319,* 1238–1240.
4. Esswein, M. J.; Nocera, D. G. *Chem. Rev.* **2007,** *107,* 4022–4047.
5. *Basic Research Needs for the Hydrogen Economy: Report on the Basic Energy Science Workshop on Hydrogen Production, Storage and Use, BES, US DOE, May 13-15, 2003,* URL http://www.sc.doe.gov/bes/hydrogen.pdf.
6. Bard, A. J.; Fox, M. A. *Acc. Chem. Res.* **1995,** *28,* 141–145.
7. Eisenberg, R.; Nocera, D. G. *Inorg. Chem.* **2005,** *44,* 6799–6801.
8. Fujishima, A.; Honda, K. *Nature* **1972,** *238,* 37.
9. Wrighton, M. S.; Ellis, A. B.; Wolczanski, P. T.; Morse, D. L.; Abrahamson, H. B.; Ginley, D. S. *J. Am. Chem. Soc.* **1976,** *98,* 2774–2779.
10. Gerischer, H. In *Solar Power and Fuels;* Bolton, J. R.; Ed.; Academic Press: NY, 1977; pp 77–112.
11. Scaife, D. E. *Solar Energy* **1980,** *25,* 41–54.
12. Torres, G. R.; Lindgren, T.; Lu, J.; Granqvist, C. G.; Lindquist, S. E. *J. Phys. Chem. B* **2004,** *108,* 5995–6003.
13. Asahi, R.; Morikawa, T.; Ohwaki, T.; Aoki, K.; Taga, Y. *Science* **2001,** *293,* 269–271.
14. Sakthivel, S.; Kisch, H. *Chemphyschem* **2003,** *4,* 487–490.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

15. Gole, J. L.; Stout, J. D.; Burda, C.; Lou, Y. B.; Chen, X. B. *J. Phys. Chem. B* **2004,** *108,* 1230–1240.
16. Nakamura, R.; Tanaka, T.; Nakato, Y. *J. Phys. Chem. B* **2004,** *108,* 10617–10620.
17. Chen, X. B.; Burda, C. *J. Phys. Chem. B* **2004,** *108,* 15446–15449.
18. Livraghi, S.; Paganini, M. C.; Giamello, E.; Selloni, A.; Di Valentin, C.; Pacchioni, G. *J. Am. Chem. Soc.* **2006,** *128,* 15666–15671.
19. Balcerski, W.; Ryu, S. Y.; Hoffmann, M. R. *J. Phys. Chem. C* **2007,** *111,* 15357–15362.
20. Kisch, H.; Sakthivel, S.; Janczarek, M.; Mitoraj, D. *J. Phys. Chem. C* **2007,** *111,* 11445–11449.
21. Nakano, Y.; Morikawa, T.; Ohwaki, T.; Taga, Y. *Physica B-Condensed Matter* **2006,** *376,* 823–826.
22. Khan, S. U. M.; Al-Shahry, M.; Ingler, W. B. *Science* **2002,** *297,* 2243–2245.
23. Tang, J.; Wu, Y. Y.; McFarland, E. W.; Stucky, G. D. *J. Chem. Soc. Chem. Commun.* **2004,** 1670–1671.
24. Nagaveni, K.; Hegde, M. S.; Ravishankar, N.; Subbanna, G. N.; Madras, G. *Langmuir* **2004,** *20,* 2900–2907.
25. Sakthivel, S.; Kisch, H. *Angew. Chem. Int. Ed.* **2003,** *42,* 4908–4911.
26. Zhao, W.; Ma, W. H.; Chen, C. C.; Zhao, J. C.; Shuai, Z. G. *J. Am. Chem. Soc.* **2004,** *126,* 4782–4783.
27. Lu, N.; Quan, X.; Li, J. Y.; Chen, S.; Yu, H. T.; Chen, G. H. *J. Phys. Chem. C* **2007,** *111,* 11836–11842.
28. Su, Y. L.; Zhang, X. W.; Han, S.; Chen, X. Q.; Lei, L. C. *Electrochem. Commun.* **2007,** *9,* 2291–2298.
29. Maeda, K.; Domen, K. *J. Phys. Chem. C* **2007,** *111,* 7851–7861.
30. Kasahara, A.; Nukumizu, K.; Hitoki, G.; Takata, T.; Kondo, J. N.; Hara, M.; Kobayashi, H.; Domen, K. *J. Phys. Chem. A* **2002,** *106,* 6750–6753.
31. Hitoki, G.; Takata, T.; Kondo, J. N.; Hara, M.; Kobayashi, H.; Domen, K. *J. Chem. Soc. Chem. Commun.* **2002,** 1698–1699.
32. Hara, M.; Takata, T.; Kondo, J. N.; Domen, K. *Catal. Today* **2004,** *90,* 313–317.
33. Hisatomi, T.; Hasegawa, K.; Teramura, K.; Takata, T.; Hara, M.; Domen, K. *Chem. Lett.* **2007,** *36,* 558–559.
34. Lee, Y.; Terashima, H.; Shimodaira, Y.; Teramura, K.; Hara, M.; Kobayashi, H.; Domen, K.; Yashima, M. *J. Phys. Chem. C* **2007,** *111,* 1042–1048.
35. Maeda, K.; Teramura, K.; Saito, N.; Inoue, Y.; Kobayashi, H.; Domen, K. *Pure Appl. Chem.* **2006,** *78,* 2267–2276.
36. Sun, X.; Maeda, K.; Le Faucheur, M.; Teramura, K.; Domen, K. *Appl. Catal. A* **2007,** *327,* 114–121.
37. Yashima, M.; Maeda, K.; Teramura, K.; Takata, T.; Domen, K. *Chem. Phys. Lett.* **2005,** *416,* 225–228.
38. Maeda, K.; Teramura, K.; Lu, D. L.; Takata, T.; Saito, N.; Inoue, Y.; Domen, K. *Nature* **2006,** *440,* 295.
39. Yashima, M.; Maeda, K.; Teramura, K.; Takata, T.; Domen, K. *Mater. Trans.* **2006,** *47,* 295–297.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

40. Maeda, K.; Teramura, K.; Masuda, H.; Takata, T.; Saito, N.; Inoue, Y.; Domen, K. *J. Phys. Chem. B* **2006,** *110,* 13107–13112.
41. Maeda, K.; Teramura, K.; Lu, D. L.; Takata, T.; Saito, N.; Inoue, Y.; Domen, K. *J. Phys. Chem. B* **2006,** *110,* 13753–13758.
42. Maeda, K.; Teramura, K.; Saito, N.; Inoue, Y.; Domen, K. *J. Catal.* **2006,** *243,* 303–308.
43. Maeda, K.; Teramura, K.; Lu, D. L.; Saito, N.; Inoue, Y.; Domen, K. *J. Phys. Chem. C* **2007,** *111,* 7554–7560.
44. Teramura, K.; Maeda, K.; Saito, T.; Takata, T.; Saito, N.; Inoue, Y.; Domen, K. *J. Phys. Chem. B* **2005,** *109,* 21915–21921.
45. Ishikawa, A.; Takata, T.; Kondo, J. N.; Hara, M.; Kobayashi, H.; Domen, K. *J. Am. Chem. Soc.* **2002,** *124,* 13547–13553.
46. Ishikawa, A.; Takata, T.; Matsumura, T.; Kondo, J. N.; Hara, M.; Kobayashi, H.; Domen, K. *J. Phys. Chem. B* **2004,** *108,* 2637–2642.
47. Ogisu, K.; Ishikawa, A.; Teramura, K.; Toda, K.; Hara, M.; Domen, K. *Chem. Lett.* **2007,** *36,* 854–855.
48. Mrowetz, M.; Balcerski, W.; Colussi, A. J.; Hoffman, M. R. *J. Phys. Chem. B* **2004,** *108,* 17269–17273.
49. Fujishima, A. *Science* **2003,** *301,* 5640.
50. Lackner, K. S. *Science* **2003,** *301,* 5640.
51. Hagglund, C.; Gratzel, M.; Kasemo, B. *Science* **2003,** *301,* 5640.
52. Park, J. H.; Kim, S.; Bard, A. J. *Nano Lett.* **2006,** *6,* 24–28.
53. Jensen, L. L.; Muckerman, J. T.; Newton, M. D. *J. Phys. Chem. C* **2008,** *112*, 3439–3446.
54. Ferreira, K. N.; Iverson, T. M.; Maghlaoui, K.; Barber, J.; Iwata, S. *Science* **2004,** *303,* 1831–1838.
55. Ruttinger, W.; Dismukes, G. C. *Chem. Rev.* **1997,** *97,* 1–24.
56. Hagfeldt, A.; Gratzel, M. *Chem. Rev.* **1995,** *95,* 49–68.
57. Gratzel, M. *Nature* **2001,** *414,* 338–344.
58. Bak, T.; Nowotny, J.; Rekas, M.; Sorrell, C. C. *Int. J. Hydrogen Energy* **2002,** *27,* 991–1022.
59. Wrighton, M. S.; Wolczanski, P. T.; Ellis, A. B. *J. Solid State Chem.* **1977,** *22,* 17–29.
60. Nozik, A. J.; Memming, R. *J. Phys. Chem. B* **1996,** *100,* 13061–13078.
61. Linsebigler, A. L.; Lu, G. Q.; Yates, J. T. *Chem. Rev.* **1995,** *95,* 735–758.
62. Yoshida, Y.; Matsuoka, M.; Moon, S. C.; Mametsuka, H.; Suzuki, E.; Anpo, M. *Res. Chem. Intermed.* **2000,** *26,* 567–574.
63. Lehn, J. M.; Sauvage, J. P.; Ziessel, R. *Nouv. J. Chim.* **1980,** *4,* 623–627.
64. Domen, K.; Kudo, A.; Onishi, T.; Kosugi, N.; Kuroda, H. *J. Phys. Chem.* **1986,** *90,* 292–295.
65. Domen, K.; Kudo, A.; Onishi, T. *J. Catal.* **1986,** *102,* 92–98.
66. Yashima, M.; Maeda, K.; Teramura, K.; Takata, T.; Domen, K. *Chem. Phys. Lett.* **2005,** *416,* 225–228.
67. Nambu, A.; Graciani, J.; Rodriguez, J. A.; Wu, Q.; Fujita, E.; Sanz, J. F. *J. Chem. Phys.* **2006,** *125,* 094706.
68. Cheung, S. H.; Nachimuthu, P.; Joly, A. G.; Engelhard, M. H.; Bowman, M. K.; Chambers, S. A. *Surf. Sci.* **2007,** *601,* 1754–1762.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

69. Shao, R.; Wang, C. M.; McCready, D. E.; Droubay, T. C.; Chambers, S. A. *Surf. Sci.* **2007,** *601,* 1582–1589.
70. Chen, H. Y.; Nambu, A.; Wen, W.; Graciani, J.; Zhong, Z.; Hanson, J. C.; Fujita, E.; Rodriguez, J. A. *J. Phys. Chem. C* **2007,** *111,* 1366–1372.
71. Maeda, K.; Teramura, K.; Takata, T.; Hara, M.; Saito, N.; Toda, K.; Inoue, Y.; Kobayashi, H.; Domen, K. *J. Phys. Chem. B* **2005,** *109,* 20504–20510.
72. Maeda, K.; Takata, T.; Hara, M.; Saito, N.; Inoue, Y.; Kobayashi, H.; Domen, K. *J. Am. Chem. Soc.* **2005,** *127,* 8286–8287.
73. Wei, S. H.; Ferreira, L. G.; Bernard, J. E.; Zunger, A. *Phys. Rev. B* **1990,** *42,* 9622–9649.
74. Lichtenstein, A. I.; Anisimov, V. I.; Zaanen, J. *Phys. Rev. B* **1995,** *52,* R5467–R5470.
75. Anisimov, V. I.; Aryasetiawan, F.; Lichtenstein, A. I. *J. Phys.: Condens. Matter* **1997,** *9,* 767–808.
76. Erhart, P.; Albe, K.; Klein, A. *Phys. Rev. B* **2006,** *73,* 205203.
77. Wei, S. H.; Zunger, A. *J. Appl. Phys.* **1995,** *78,* 3846–3856.
78. Wei, S. H.; Zunger, A. *Phys. Rev. Lett.* **1996,** *76,* 664–667.
79. Chen, H.; Wen, W.; Wang, Q.; Hanson, J. C.; Muckerman, J. T.; Fujita, E.; Frenkel, A.; Rodriguez, J. A. *J. Phys. Chem. C*, submitted.
80. Meyer, T. J.; Huynh, M. H. V. *Inorg. Chem.* **2003,** *42,* 8140–8160.
81. Yagi, M.; Kaneko, M. *Chem. Rev.* **2001,** *101,* 21–35.
82. Morris, N. D.; Suzuki, M.; Mallouk, T. E. *J. Phys. Chem. A* **2004,** *108,* 9115–9119.
83. Wada, T.; Tsuge, K.; Tanaka, K. *Inorg. Chem.* **2001,** *40,* 329–337.
84. Wada, T.; Tsuge, K.; Tanaka, K. *Angew. Chem. Int. Ed.* **2000,** *39,* 1479.
85. Pecoraro, V. L.; Baldwin, M. J.; Gelasco, A. *Chem. Rev.* **1994,** *94,* 807–826.
86. Gersten, S. W.; Samuels, G. J.; Meyer, T. J. *J. Am. Chem. Soc.* **1982,** *104,* 4029–4030.
87. Gilbert, J. A.; Eggleston, D. S.; Murphy, W. R.; Geselowitz, D. A.; Gersten, S. W.; Hodgson, D. J.; Meyer, T. J. *J. Am. Chem. Soc.* **1985,** *107,* 3855–3864.
88. Vining, W. J.; Meyer, T. J. *Inorg. Chem.* **1986,** *25,* 2023–2033.
89. Chronister, C. W.; Binstead, R. A.; Ni, J. F.; Meyer, T. J. *Inorg. Chem.* **1997,** *36,* 3814–3815.
90. Lebeau, E. L.; Adeyemi, S. A.; Meyer, T. J. *Inorg. Chem.* **1998,** *37,* 6476–6484.
91. Binstead, R. A.; Chronister, C. W.; Ni, J. F.; Hartshorn, C. M.; Meyer, T. J. *J. Am. Chem. Soc.* **2000,** *122,* 8464–8473.
92. Bartolotti, L. J.; Pedersen, L. G.; Meyer, T. J. *Int. J. Quantum Chem.* **2001,** *83,* 143–149.
93. Hurst, J. K. *Coord. Chem. Rev.* **2005,** *249,* 313–328.
94. Hurst, J. K.; Zhou, J. Z.; Lei, Y. B. *Inorg. Chem.* **1992,** *31,* 1010–1017.
95. Yang, X; Baik, M. H. *J. Am. Chem. Soc.* **2006,** *128*, 7476–7485.
96. Lei, Y. B.; Hurst, J. K. *Inorg. Chem.* **1994,** *33,* 4460–4467.
97. Yamada, H.; Hurst, J. K. *J. Am. Chem. Soc.* **2000,** *122,* 5303–5311.
98. Yamada, H.; Koike, T.; Hurst, J. K. *J. Am. Chem. Soc.* **2001,** *123,* 12775–12780.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

99. Yamada, H.; Siems, W. F.; Koike, T.; Hurst, J. K. *J. Am. Chem. Soc.* **2004,** *126,* 9786–9795.
100. Limburg, J.; Vrettos, J. S.; Liable-Sands, L. M.; Rheingold, A. L.; Crabtree, R. H.; Brudvig, G. W. *Science* **1999,** *283,* 1524–1527.
101. Limburg, J.; Crabtree, R. H.; Brudvig, G. W. *Inorg. Chim. Acta* **2000,** *297,* 301–306.
102. Limburg, J.; Vrettos, J. S.; Chen, H. Y.; de Paula, J. C.; Crabtree, R. H.; Brudvig, G. W. *J. Am. Chem. Soc.* **2001,** *123,* 423–430.
103. Ruettinger, W.; Yagi, M.; Wolf, K.; Bernasek, S.; Dismukes, G. C. *J. Am. Chem. Soc.* **2000,** *122,* 10353–10357.
104. Yagi, M.; Osawa, Y.; Sukegawa, N.; Kaneko, M. *Langmuir* **1999,** *15,* 7406–7408.
105. Yagi, M.; Tokita, S.; Nagoshi, K.; Ogino, I.; Kaneko, M. *J. Chem. Soc. Faraday Trans.* **1996,** *92,* 2457–2461.
106. Yagi, M.; Nagoshi, K.; Kaneko, M. *J. Phys. Chem. B* **1997,** *101,* 5143–5146.
107. Comte, P.; Nazeeruddin, M. K.; Rotzinger, F. P.; Frank, A. J.; Gratzel, M. *J. Mol. Catal.* **1989,** *52,* 63–84.
108. Yagi, M.; Narita, K. *J. Am. Chem. Soc.* **2004,** *126,* 8084–8085.
109. Shimazaki, Y.; Nagano, T.; Takesue, H.; Ye, B. H.; Tani, F.; Naruta, Y. *Angew. Chem. Int. Ed.* **2004,** *43,* 98–100.
110. Sens, C.; Romero, I.; Rodriguez, M.; Llobet, A.; Parella, T.; Benet-Buchholz, J. *J. Am. Chem. Soc.* **2004,** *126,* 7798–7799.
111. Zong, R.; Thummel, R. P. *J. Am. Chem. Soc.* **2005,** *127,* 12802–12803.
112. Nagoshi, K.; Yamashita, S.; Yagi, M.; Kaneko, M. *J. Mol. Catal. A: Chem.* **1999,** *144,* 71–76.
113. Liu, F.; Concepcion, J. J.; Jurss, J. W.; Cardolaccia, T.; Templeton, J. L.; Meyer, T. J. *Inorg. Chem.* **2008,** *47,* 1727–1752.
114. Hurst, J. K.; Cape, J. L.; Clark, A. E.; Das, S.; Qin, C. *Inorg. Chem.* **2008,** *47,* 1753–1764.
115. Cape, J. L.; Hurst, J. K. *J. Am. Chem. Soc.* **2008,** *130,* 827–829.
116. Muckerman, J. T.; Polyansky, D. E.; Wada, T.; Tanaka, K.; Fujita, E. *Inorg. Chem.* **2008,** *47,* 1787–1802.
117. Kobayashi, K.; Ohtsu, H.; Wada, T.; Kato, T.; Tanaka, K. *J. Am. Chem. Soc.* **2003,** *125,* 6729–6739.
118. Haga, M.; Isobe, K.; Boone, S. R.; Pierpont, C. G. *Inorg. Chem.* **1990,** *29,* 3795–3799.
119. Pierpont, C. G.; Buchanan, R. M. *Coord. Chem. Rev.* **1981,** *38,* 45–87.
120. Pierpont, C. G.; Lange, C. W., *Progress in Inorganic Chemistry,* **1994,** *41,* 331–442.
121. Auburn, P. R.; Dodsworth, E. S.; Haga, M.; Liu, W.; Nevin, W. A.; Lever, A. B. P. *Inorg. Chem.* **1991,** *30,* 3502–3512.
122. Lever, A. B. P.; Auburn, P. R.; Dodsworth, E. S.; Haga, M.; Wei, L.; Melnik, M.; Nevin, W. A. *J. Am. Chem. Soc.* **1988,** *110,* 8076–8084.
123. Haga, M.; Dodsworth, E. S.; Lever, A. B. P.; Boone, S. R.; Pierpont, C. G. *J. Am. Chem. Soc.* **1986,** *108,* 7413–7414.
124. Bhattacharya, S.; Boone, S. R.; Fox, G. A.; Pierpont, C. G. *J. Am. Chem. Soc.* **1990,** *112,* 1088–1096.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

125. Bhattacharya, S.; Pierpont, C. G. *Inorg. Chem.* **1992,** *31,* 35–39.
126. Masui, H.; Lever, A. B. P.; Auburn, P. R. *Inorg. Chem.* **1991,** *30,* 2402–2410.
127. Takeuchi, K. J.; Thompson, M. S.; Pipes, D. W.; Meyer, T. J. *Inorg. Chem.* **1984,** *23,* 1845–1851.
128. Gloaguen, F.; Lawrence, J. D.; Rauchfuss, T. B. *J. Am. Chem. Soc.* **2001,** *123,* 9476–9477.
129. Justice, A. K.; Linck, R. C.; Rauchfuss, T. B.; Wilson, S. R. *J. Am. Chem. Soc.* **2004,** *126,* 13214–13215.
130. van der Vlugt, J. I.; Rauchfuss, T. B.; Whaley, C. M.; Wilson, S. R. *J. Am. Chem. Soc.* **2005,** *127,* 16012–16013.
131. Lyon, E. J.; Georgakaki, I. P.; Reibenspies, J. H.; Darensbourg, M. Y. *J. Am. Chem. Soc.* **2001,** *123,* 3268–3278.
132. Zhao, X.; Georgakaki, I. P.; Miller, M. L.; Yarbrough, J. C.; Darensbourg, M. Y. *J. Am. Chem. Soc.* **2001,** *123,* 9710–9711.
133. Mejia-Rodriguez, R.; Chong, D.; Reibenspies, J. H.; Soriaga, M. P.; Darensbourg, M. Y. *J. Am. Chem. Soc.* **2004,** *126,* 12004–12014.
134. Evans, D. J.; Pickett, C. J. *Chem. Soc. Rev.* **2003,** *32,* 268–275.
135. Tard, C.; Liu, X.; Ibrahim, S. K.; Bruschi, M.; Gioia, L. D.; Davies, S.; Yang, X.; Wang, L.-S.; Sawers, G.; Pickett, C. J. *Nature* **2005,** *434,* 610–613.
136. Liu, X.; Ibrahim, S. K.; Tard, C.; Pickett, C. J. *Coord. Chem. Rev.* **2005,** *249,* 1641–1652.
137. Borg, S. J.; Behrsing, T.; Best, S. P.; Razavet, M.; Liu, X.; Pickett, C. J. *J. Am. Chem. Soc.* **2004,** *126,* 16988–16999.
138. Schwartz, L.; Eilers, G.; Eriksson, L.; Gogoll, A.; Lomoth, R.; Ott, S. *J. Chem. Soc. Chem. Commun.* **2006,** 520–522.
139. Ott, S. K., M.; Åkermark, B.; Sun, L.; Lomoth, R. *Angew. Chem. Int. Ed* **2004,** *43,* 1006–1009.
140. Curtis, C. J.; Miedaner, A.; Ciancanelli, R.; Ellis, W. W.; Noll, B. C.; DuBois, M. R.; DuBois, D. L. *Inorg. Chem.* **2003,** *42,* 216–227.
141. Wilson, A. D.; Newell, R. H.; McNevin, M. J.; Muckerman, J. T.; DuBois, M. R.; DuBois, D. L. *J. Am. Chem. Soc.* **2006,** *128,* 358–366.
142. Henry, R. M.; Shoemaker, R. K.; DuBois, D. L.; DuBois, M. R. *J. Am. Chem. Soc.* **2006,** *128,* 3002–3010.
143. Wilson, A. D.; Shoemaker, R. K.; Miedaner, A.; Muckerman, J. T.; DuBois, D. L.; DuBois, M. R. *Proc. Natl. Acad. Sci. USA* **2007,** *104,* 6951–6956.
144. Hu, X. L.; Cossairt, B. M.; Brunschwig, B. S.; Lewis, N. S.; Peters, J. C. *J. Chem. Soc. Chem. Commun.* **2005,** 4723–4725.
145. Hu, X.; Brunschwig, B. S.; Peters, J. C. *J. Am. Chem. Soc.* **2007,** *129,* 8988–8998.
146. Connolly, P.; Espenson, J. H. *Inorg. Chem.* **1986,** *25,* 2684–2688.
147. Matsuoka, S. Y., K.; Ogata, T.; Kusaba, M.; Nakashima, N.; Fujita, E.; Yanagida, S. *J. Am. Chem. Soc.* **1993,** *115,* 601–609.
148. Ogata, T.; Yanagida, S.; Brunschwig, B. S.; Fujita, E. *J. Am. Chem. Soc.* **1995,** *117,* 6708–6716.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

149. Fujita, E.; Creutz, C.; Sutin, N.; Szalda, D. J. *J. Am. Chem. Soc.* **1991,** *113,* 343–353.
150. Creutz, C.; Schwarz, H. A.; Wishart, J. F.; Fujita, E.; Sutin, N. *J. Am. Chem. Soc.* **1991,** *113,* 3361–3371.
151. Fujita, E.; Creutz, C.; Sutin, N.; Brunschwig, B. S. *Inorg. Chem.* **1993,** *32,* 2657–2662.
152. Fujita, E.; Haff, J.; Sanzenbacher, R.; Elias, H. *Inorg. Chem.* **1994,** *33,* 4627–4628.
153. Sutin, N.; Creutz, C.; Fujita, E. *Comments on Inorganic Chemistry* **1997,** *19,* 67–92.
154. Dhanasekaran, T.; Grodkowski, J.; Neta, P.; Hambright, P.; Fujita, E. *J. Phys. Chem. A* **1999,** *103,* 7742–7748.
155. Behar, D.; Dhanasekaran, T.; Neta, P.; Hosten, C. M.; Ejeh, D.; Hambright, P.; Fujita, E. *J. Phys. Chem. A* **1998,** *102,* 2870–2877.
156. Fujita, E. *Coord. Chem. Rev.* **1999,** *186,* 373–384 and references therein.
157. Grodkowski, J.; Dhanasekaran, T.; Neta, P.; Hambright, P.; Brunschwig, B. S.; Shinozaki, K.; Fujita, E. *J. Phys. Chem. A* **2000,** *104,* 11332–11339.
158. Fujita, E.; Brunschwig, B. S. In *Electron Transfer in Chemistry;* Balzani, V.; Ed.; Wiley-VCH: Weinheim, Germany, 2001; pp 88–126.
159. Grodkowski, J.; Neta, P.; Fujita, E.; Mahammed, A.; Simkhovich, L.; Gross, Z. *J. Phys. Chem. A* **2002,** *106,* 4772–4778.
160. Fisher, B.; Eisenberg, R. *J. Am. Chem. Soc.* **1980,** *102,* 7361–7365.
161. Lehn, J.-M.; Ziessel, R. *Proc. Natl. Acad. Sci. USA* **1982,** *79,* 701–704.
162. Ziessel, R.; Hawecker, J.; Lehn, J.-M. *Helv. Chim. Acta* **1986,** *69,* 1065–1084.
163. Kitamura, N.; Tazuke, S. *Chem. Lett.* **1983,** 1109–1112.
164. Willner, I.; Maidan, R.; Mandler, D.; Duerr, H.; Doerr, G.; Zengerle, K. *J. Am. Chem. Soc.* **1987,** *109,* 6080–6086.
165. Maidan, R.; Willner, I. *J. Am. Chem. Soc.* **1986,** *108,* 8100–8101.
166. Ishida, H.; Tanaka, K.; Tanaka, T. *Organometallics* **1987,** *6,* 181–186.
167. Ishida, H.; Terada, T.; Tanaka, K.; Tanaka, T. *Inorg. Chem.* **1990,** *29,* 905–911.
168. Tinnemans, A. T. A.; Koster, T. P. M.; Thewissen, D. H. M. W.; Mackor, A. *Recl. Trav. Chim. Pays. -Bas* **1984,** *103,* 288–295.
169. Grant, J. L.; Goswami, K.; Spreer, L. O.; Otvos, J. W.; Calvin, M. *J. Chem. Soc. Dalton Trans* **1987,** 2105-2109.
170. Kimura, E.; Wada, S.; Shionoya, M.; Okazaki, Y. *Inorg. Chem.* **1994,** *33,* 770–778.
171. Takeda, H.; Koike, K.; Inoue, H.; Ishitani, O. *J. Am. Chem. Soc.* **2008,** *130,* 2023–2031.
172. Sato, S.; Koike, K.; Inoue, H.; Ishitani, O. *Photochem. Photobiol. Sci.* **2007,** *6,* 454–461.
173. Tsubaki, H.; Sekine, A.; Ohashi, Y.; Koike, K.; Takeda, H.; Ishitani, O. *J. Am. Chem. Soc.* **2005,** *44,* 15544–15555.
174. Tsubaki, H.; Sugawara, A.; Takeda, H.; Gholamkhass, B.; Koike, K.; Ishitani, O. *Res. Chem. Intermed.* **2007,** *33,* 37–48.
175. Gholamkhass, B.; Mametsuka, H.; Koike, K.; Tanabe, T.; Furue, M.; Ishitani, O. *Inorg. Chem.* **2005,** *44,* 2326–2336.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

176. Hori, H.; Ishihara, J.; Koike, K.; Takeuchi, K. J.; Ibusuki, T.; Ishitani, O. *J. Photochem. Photobiol. A* **1999,** *120,* 119–124.
177. Konno, H.; Kobayashi, A.; Sakamoto, K.; Fagalde, F.; Katz, N. E.; Saitoh, H.; Ishitani, O. *Inorg. Chim. Acta* **2000,** *299,* 155–163.
178. Koike, K.; Hori, H.; Ishizuka, M.; Westwell, J. R.; Takeuchi, K. J.; Ibusuki, T.; Enjouji, K.; Konno, H.; Sakamoto, K.; Ishitani, O. *Organometallics* **1997,** *16,* 5724–5729.
179. Ishitani, O.; George, M. W.; Ibusuki, T.; Johnson, F. P. A.; Koike, K.; Nozaki, K.; Pac, C. J.; Turner, J. J.; Westwell, J. R. *Inorg. Chem.* **1994,** *37,* 4712–4717.
180. Pugh, J. R.; Bruce, M. R. M.; Sullivan, B. P.; Meyer, T. J. *Inorg. Chem.* **1991,** *30,* 86–91.
181. Otoole, T. R.; Sullivan, B. P.; Bruce, M. R. M.; Margerum, L. D.; Murray, R. W.; Meyer, T. J. *J. Electroanal. Chem.* **1989,** *259,* 217–239.
182. Bruce, M. R. M.; Megehee, E.; Sullivan, B. P.; Thorp, H. H.; Otoole, T. R.; Downard, A.; Pugh, J. R.; Meyer, T. J. *Inorg. Chem.* **1992,** *31,* 4864–4873.
183. Nagao, H.; Mizukawa, T.; Tanaka, K. *Inorg. Chem.* **1994,** *33,* 3415–3420.
184. Tanaka, K.; Ooyama, D. *Coord. Chem. Rev.* **2002,** *226,* 211–218.
185. Sweet, J. R.; Graham, W. A. *J. Am. Chem. Soc.* **1982,** *104,* 2811–2815.
186. Koizumi, T.; Tanaka, K. *Angew. Chem. Int. Ed.* **2005,** *44,* 5891-5894.
187. Polyansky, D. E.; Cabelli, D.; Muckerman, J. T.; Fujita, E.; Koizumi, T.; Fukushima, T.; Wada, T.; Tanaka, K. *Angew. Chem. Int. Ed.* **2007,** *46,* 4169–4172.
188. Polyansky, D. E.; Cabelli, D.; Muckerman, J. T.; Fukushima, T.; Tanaka, K.; Fujita, E. *Inorg. Chem.* **2008,** *47,* 3958–3968.

Downloaded by COLUMBIA UNIV on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch015

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

Chapter 16

Chemical Evolution and Biomimetic Chemistry

Christopher J. Bender

Department of Chemistry, Fordham University, 441 East Fordham Road, Bronx, NY 10458

Biological materials and processes inspire many engineered materials and devices because of their specialized nature and oftentimes high efficiency. Although optimized *in vivo* through hereditary evolution, primordial versions of biological materials must have evolved *via* chemical routes, and therefore it stands to reason that biomimetic chemistry is complementary to research in chemical evolution. The types of reactions described by Miller-Urey, for example, represent synthetic routes from inorganic gas mixtures to organic molecules that, in turn, serve as a pool of precursors in combinatorial polymerization reactions. On a primordial Earth, these reactions would presumably be subject to a geochemical form of molecular imprinting, which may justify an origin of peptide motifs and enzyme archetypes that pre-existed hereditary mechanisms of optimization.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Introduction

Since it was first discovered that biologically-derived materials could be used to enhance industrial chemical reactions, chemists have been motivated to replicate the action of these biological agents by synthetic means, either to produce the activating agent more cheaply (and in greater quantity), or to improve upon the process. The 'ferments', as some of these agents were called, were recognized as being a unique type of catalyst, and early attempts to synthetically re-create their action proceeded *via* a merger of a general catalyst with a supporting material that imposed some sort of spatial constraint upon the reactants, for example, colloidal metal (*1*). But basic research into the chemical behavior of the biological catalysts soon established that they were unique from the so-called general catalysts: they were true macro-molecules, as opposed to colloids, comprised of amino acid chains assembled by an amide linkage, and both the type of amino acid and their order in the macromolecular chain determined structure and function. The site of the chemical reaction, whether acting as a receptor for information transfer or as a locale for catalytic reaction, featured a specific order of the atomic structure that, in turn, enabled the active site to selectively bind its targeted substrate. And as chemical means of analyzing the active site gave way to structural analysis of proteins in their entirety, one increasingly finds the correlation between protein structure and function described in terminology that reflects the dynamic behavior of proteins. Proteins are often described as molecular machines and devices, and the protein as device has in fact entered the popular literature (*2*,*3*).

Biomimetic chemistry is a term that was introduced in the 1970s (*4*-*6*) and may be loosely defined as the synthetic replication of chemistries associated with biological materials. It is meant to describe an experimental approach to the chemical modeling of biomaterials that differs from classical bio-organic chemistry in the sense that biomimetic models are 'thematic', in other words, function-driven, rather than a replication of the biological original. For example, many early bio-organic models would attempt to synthesize small-molecule versions of an enzyme active site (i.e. the active site without the protein support), whereas biomimetic chemistry takes the viewpoint that the enzyme active site is simply a functionalized pocket, and that the pocket's function can be replicated by some general molecular cavity, such as afforded by cyclodextrin. And since the biomimetic emphasis is on function rather than explicit replication of biological structure, very novel approaches to the problem of 'molecule as device' have been undertaken, with explorations of novel synthetic stratagems such as templating, self-assembly, etc., all having a biological exemplar.

At its origin, life, or perhaps more accurately, 'living chemistry', presumably had some of the chemical attributes that we recognize as being unique and desirous to replicate. It is therefore appropriate to ask from what types of chemical dynamics do the archetypes of "Nature's Robots" (*3*) come, and whether it is possible to synthesize models of proto-enzymes *via* simple reactions using directed-synthesis methods; in a sense, geochemical biomimicry. Chemical evolution denotes the processes by which simple inorganic molecules such as the volatiles found in planetary atmospheres react and combine to form

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

small molecules of low-volatility, which then serve as building blocks for large non-volatile and complicated molecules (*7,8*). And so explorations of plausible routes of chemical evolution and synthetic biomimetic chemistry are complementary, since both endeavor to find *de novo* methods of synthesizing materials that mimic the chemistry of living matter.

The subject of this chapter concerns the chemical evolution of one type of biological polymer, namely the proteins, which are synthesized *in vivo* from a finite set of monomers *via* a collection of interacting chemical reactions broadly defined as metabolism. The sequence of the monomers in the polymer imparts information – polymer structure and function – that is derived from a template within the synthetic apparatus of the cell. But while the chemical linkage of amino acids to generate a polypeptide chain is simple, imparting information to that chain is not. Chemical information is represented by atomic order and may be translated *via* templated reactions, in which a primary source of information (the template molecule) determines the assembly of a complementary molecule (*9*) for example, the translation of a DNA sequence into a specific protein *in vivo*. But it stands to reason that at some point during the earliest stages of chemical evolution on the primordial Earth there was no coupling between the polynucleic acids, the so-called 'RNA World' with its own polynucleic acid catalysts (*10*), and the polyamino acids. And so, one may argue that cells equipped with a polymer-to-polymer templating machinery (i.e. the 'progenotes', reference *11*) evolved from an ancestor (or ancestors) that may be described as a viscous electrolyte solution containing small organic molecules and catalytic macromolecules, which may or may not be encapsulated. In other words, we are searching for reactions in complex mixtures of simple organic molecules that exhibit some manner of templating and so yield a set of, in effect, supported-metal catalysts that range in function in much the same way as the various classes of enzymes. Such reactions would, in turn, act reflexively upon the original pool of small organic molecules and set in motion what we call metabolism.

Chemical Evolution as a Strategy for Biomimetic Synthesis

Anabolic metabolism denotes those reactions of the cell that build complicated molecules from simple precursors that are derived from the degradative reactions of catabolic metabolism. The cell, in other words, may be regarded as a dense pool of small organic molecular building-blocks that might be denoted as 'synthons', to borrow terminology from modern approaches to synthetic organic chemistry (*12-14*). For example, Calvin's textbook (*7*) contains introductory chapters that describe what he calls 'molecular paleontology', which resembles a retrosynthetic analysis of intermediary metabolites. The reactions that we collectively call 'metabolism' therefore correspond to the multiple dynamic processes that occur within this pool of synthons, which act in an integrated fashion with both negative and positive feedback and ultimately determine the compositional make-up of the system at any point in time.

Chemical evolution presumes that there was a primordial source of these synthons that comprise the basis of anabolic metabolism. Molecules exist only at

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

low temperature (a relative term), making their appearance first as diatomic molecules among late type stars with temperatures on the order of 10,000K (*15*). Representative molecules of stellar origin that are suggestive of organic molecule precursors include the diatomics CH, CN, CO, and polyatomics such as HCN, HC_nN (n = 2, odd numbers up to 13), HCO, CH_4, C_nH and C_nH_2 (n = 2, 3, 4), C_2H_4, C_2H_3N. NH_3, H_2O, and SO_2 complement this catalog of relevant stellar molecular building-blocks, but neither O_2 nor O_3 are detected in stellar spectra (*15*).

The interstellar space is populated by a rich variety of molecular species, alone or associated with (and in fact, synthesized upon) accretive particulates such as ice, dust and grains (*16*). Condensation processes in interstellar clouds lead to the production of hydrogen-rich molecules H_2O, NH_3, and CH_4, plus CO, CO_2, N_2, and O_2. Larger molecules include formic acid, formamide, methanol, ethanol, methyl formate, etc., and some spectral studies have indicated that polycyclic aromatic hydrocarbons are synthesized on grains (*16*). The basic synthons of (bio)organic synthesis are therefore found in the interstellar dust, and the chemical flux among these molecular species is presumably maintained by high energy stellar radiation.

Terrestrial planets form by accretion of particulates (planetesimals), and so the various organic molecules of abiotic origin are potentially present during planetary formation (*17,18*). Those that we would identify as organic synthons of primordial life are presumably present only in trace quantities, and we look towards energetic processes that would result in their production *via* the more abundant volatiles, namely the elements of the Archean atmosphere (*19-21*). Spark discharges in simple gas mixtures that mimic primitive planetary atmospheres yield a condensate of small organic molecules (*22,23*). This procedure has been replicated (cf. references *7,8*) using variants of the model primordial atmosphere (i.e. gas composition) and the driving force (i.e. heat, shock wave, ionizing radiation, etc.). Collectively, these experiments demonstrate that a simple combination of small inorganic molecules commonly found in planetary atmospheres will, upon an input of energy (independent of form), yield a collection of small organic molecules whose identity is recognizable as coinciding with some of the intermediates of metabolism, particularly those of methanogenic bacteria (*24*). The experiments therefore represent a region of overlap between the abiotic and biotic chemistries.

The spark discharge experiments and their ilk jibed with early hypotheses that life evolved from a primordial ocean rich in organic molecules *via*, for example, evaporative concentration (*25,26*). In this conceptual model, the pool of organic molecules containing one to four carbon atoms in their skeleton forms a synthon basis for what can be imagined as combinatorial syntheses subject to suitable reaction conditions. For example, among the organic molecules produced in the spark discharge were several of the amino acids (*22*), and because these monomeric precursors of proteins are bifunctional, they may be linked to yield polypeptides (dubbed proteinoids) *via* a thermally driven condensation reaction (*27,28*). This sequence of spark discharge, followed by thermally-driven polymerization of select monomers served as an early plausible model of a chemical evolution route from inorganic (polyatomic) molecules to macromolecules suggestive of biopolymers that was attractive because it was

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

also consistent with early geological models of chemical evolution that invoked a primordial ocean and evaporative concentration 'seaside model' of life's origins.

The conditions on the primordial Earth and the timing of life's origin have been open to debate, however, and so alternate models of chemical evolution have been described. Woese, for example, has proposed a model that reflects a primordial Earth that resembles the planet Venus in the sense that the atmosphere was predominantly CO_2 and the Earth's surface temperature was too hot for water to exist in its liquid state (*29*). The system in which chemical evolution takes place therefore consists of fine water droplets that contain organic synthons and undergo a kind of reflux between the hot terrestrial surface and the atmosphere (something akin to a sauna). As such, the small water droplets in which chemical evolution occurs are inherently "cellular" in the sense that the aerosol droplets are nearly the same size as cells and would therefore impose similar spatial constraints upon organic molecules sequestered within them. Woese also emphasizes that the large surface to volume ratio of these aerosols would render significant interfacial chemical processes. Such processes are easy to replicate in the laboratory by spraying amino acid solutions into a hot crucible (Figure 1).

Bacteria have long been known to inhabit the aerosols of clouds, and recent studies of the atmospheric aerosols have demonstrated that these aerosols are likewise rich in organic molecules (*30*). It has therefore been proposed that atmospheric aerosols comprised a medium for chemical evolution (*31,32*), although this model differs from that of Woese by invoking a role for an existing ocean (reminiscent of Oparin's primordial soup, which Woese rejects, *29*) and a photochemical reaction process. Dobson *et al.* (*31*), for example, describe a model in which the aerosol originated from a primordial ocean in the form of spray, and the aerosols, once in the atmosphere, were subjected to ultraviolet radiation, which initiated photochemical reactions. The Woese model, by contrast, largely depends on thermal processes (interaction with the Earth's hot surface) because a dense atmosphere would presumably attenuate the chemically activating ultraviolet radiation. It is also worth mentioning that the Miller-Urey experiments cited above (*22*) injected water into the model atmosphere *via* a boiling water bath, which means that these discharge reactions probably included some aerosol component.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

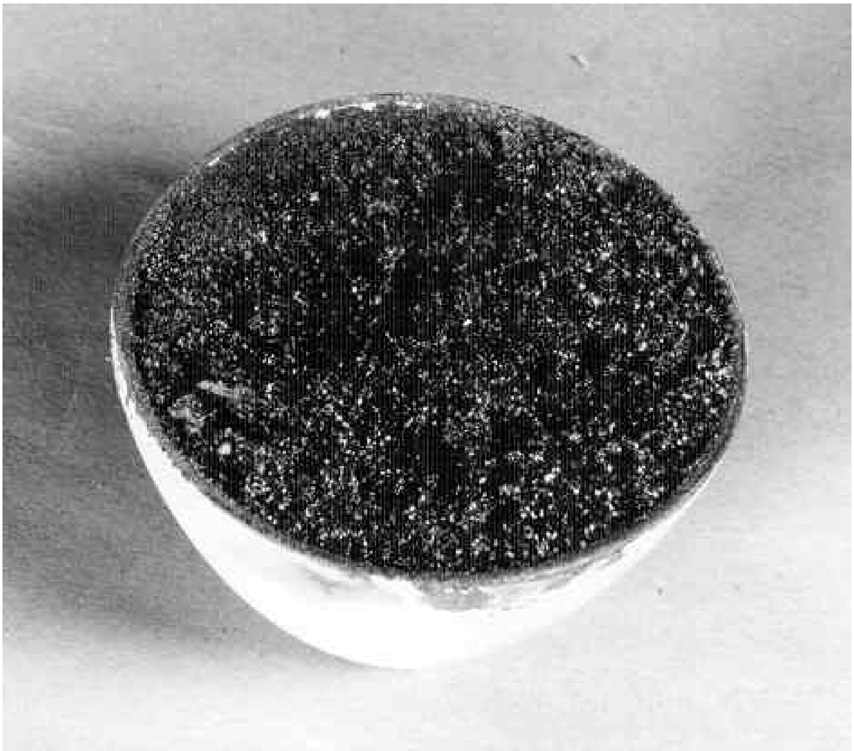


Figure 1. Proteinoid material created by spraying a solution of amino acids onto a hot crucible in a furnace.

A third model of chemical evolution has been proposed that follows from discoveries of thermophilic bacteria and organic material in the efflux of deep-sea vents (*33,34*). These submarine hydrothermal vents are recognized as gradient environments in the sense that they serve as a continuously varying physical and chemical link between the ocean and the Earth's magma and, as such, these gradient environments possess two attributes that are lacking in the aforementioned models, namely, a geochemical link and an energetic driving force that is continuous rather than intermittent. The model is also all-encompassing: heat derived from newly formed oceanic crust was the driving force for the conversion of inorganic molecules to organic synthons – the molecules of intermediary metabolism – followed by polymerization reactions in shallow oceans that are rich in these organic synthons, but there is also evidence that polymerization may have occurred at high temperature and pressure (*34*), that is, at the sea floor, despite arguments to the contrary (*35*).

The seawater at thermal vents is at high temperature and pressure, and the so-called gradient environment at the ocean floor may therefore be regarded as a supercritical fluid. Polymerization reactions can occur under such conditions, but the mechanism of chain propagation differs from that proposed in connection with the Harada-Fox model (*27*). Rather than condensation reactions, chain growth polymerization reactions under high pressure typically tend to proceed *via* free radical, ring-opening, cationic or anionic processes (cf.

references *36*,*37*). Seawater at a deep-sea thermal vent will be rich in metal cations derived from dissolved minerals, and so the conditions for polymerization are, in principle, favorable. There is, for example, evidence that long-chain hydrocarbons may be formed in hydrothermal conditions *via* Fischer-Tropsch routes (*38,39*).

Although rare in the literature, there is a precedent for a condensation-like polymerization in which the supercritical medium, in this case CO_2, facilitates the reaction to polyesters by extraction of ethylene glycol (*40*). But, apropos to abiotic synthesis of proteinoids, a condensation reaction analogous to the Fox-Harada polymerization would require water extraction. Some experiments and thermodynamic models suggest that dehydration processes are indeed possible in supercritical aqueous solutions (*41,42*), but these theoretical models must be reconciled with other experimental studies indicating that amino acids decompose *via* decarboxylation and deamination under conditions that are believed to accurately represent the gradient environment of the hydrothermal vent (*43*).

To summarize, there exist three chemical evolution models that parallel synthetic polymerization procedures and whose reaction conditions may be used as strategies for biomimetic syntheses of biopolymers from organic synthons:

1. Bulk condensation polymerization driven thermally or by an exogenous dehydration reagent, such as an acid anhydride, from a concentrated solution.
2. Emulsion polymerization from aerosols or micelles driven photochemically or thermally, but governed by the properties of interfaces and interfacial processes.
3. Polymerization in supercritical water driven thermally.

Each of these synthetic strategies has a precedent in synthetic polymer chemistry. For example, polyamides (e.g. nylon) are synthesized from melts containing their monomer; and styrene may be polymerized in micelles of laurate (*36*). High pressure polymerization is usually used to modify reaction parameters, such as rates and degree of polymerization, mechanisms of chain transfer and branching processes, and polymer properties, but the supercritical solvent enables one to optimize the solvent-reactant interaction, yielding homogeneity or inhomogeneity. The latter suggests emulsion polymerization; there is considerable overlap between the methods. But from the standpoint of synthetic biomimetic chemistry, the syntheses of C-3 and higher organic synthons from gaseous precursors and the various polymerizations of a select group of these synthons collectively suggest plausible routes from inorganic volatiles to biomimetic macromolecules in which the energy required to drive the reaction decreases as one progresses towards molecular complexity.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

Templated Reactions and the Geochemical Factor in Chemical Evolution

The three chemical evolution mechanisms of prospective proteinoid formation describe the polymerization of a select pool of monomers (i.e. amino acids) in a medium that is either homogeneous (bulk polymerization) or heterogeneous (emulsion polymerization). The thermally driven condensation polymerization of glycine (or a mixture of amino acids) from a glutamic acid melt is an example of a bulk process, and in both execution and outcome the reactions described by Harada and Fox (*27,28*) resemble the condensation polymerization of polyamides such as nylon (*44*). In general, the condensation of amides (i.e. amidation) is "rapid above 180° to 200°C, and remarkably free of side reactions" (*45*). A typical synthesis, for example, entails the heating of aminoundecanoic acid (a monomer that is a low melting liquid) at 220°C for three hours. The water that is expelled during the amide condensation is lost *via* evaporation, and any that remains within the system is immiscible with the molten mixture of growing polymer and unreacted monomer. Such amidation processes from bulk are one type of the so-called step-reactions that proceed *via* second order kinetics, and the ultimate size of the polymer is determined by the efficiency of the reaction and thus affected by experimental conditions (*45*).

Harada and Fox used pyroglutamic acid as a medium for the condensation polymerization of glycine because the latter does not melt at temperatures below 240°C (where it decomposes); at 180°C glutamic acid is converted to pyroglutamic acid *via* a condensation process that is reminiscent of Staudinger's early synthetic route towards β-lactams (*46*). The pyroglutamic acid remains liquid at temperatures above 160°C, and, as such, the condensation polymerization of glycine in a pyroglutamic acid melt is akin to the general procedure of performing reactions in fused salts. In practice, the glutamic acid may be first converted to pyroglutamic acid, and glycine then added to the melt, or the solid amino acids (glutamic acid and glycine) may be mixed prior to heating. In the latter case, glycine 'dissolves' into the pyroglutamic acid melt as it is formed.

The bulk polymerization reaction of glycine from a pyroglutamic acid melt has kinetic features that are consistent with those described by Flory (*45*) in the condensation reactions that yield polyamides; for example, a correlation between average molecular weight and the surface-volume ratio of the melt. And because it is a simple reaction that is representative of a chemical evolution step towards biological macromolecules, this thermally-driven condensation reaction may be used as a starting point from which we begin to consider the geochemical influence of the reaction. In their bulk polymerization experiments, Fox and Harada reported that the incorporation of amino acids from the melt into the proteinoid was not statistical (*28*), which suggests that some kinetic or thermodynamic factor controls the relative rate of amino acid incorporation and therefore results in a kind of selection. The result is not surprising, because this is consistent with the general features of amidation, among which there is a relationship between process efficiency and the polymer chain growth. But it does suggest that there may have been a physico-chemical mechanism for establishing amino acid motifs within proteinoids during the earliest stages of

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

chemical evolution, that is, prior to establishing a link between the polynucleic acids and polypeptides (genes and their products).

Critique of the Harda-Fox Polymerization and Refinement of Conditions

The thermally-driven condensation polymerization of amino acids, as initially described by Harada and Fox (*27,28*) and summarized in the preceding section, has been criticized on account of the apparent side reactions that occur during the melt. This criticism is summarized by Wills and Bada (*47*) in context of the history of chemical evolution research, but the gist of the problem lies with the pyrolysis of amino acids at high temperatures, which renders polymerizations more random than the desired synthetic route towards linear polypeptides.

It would appear that in their initial report (*27*), Harada and Fox were strongly influenced by concurrent discoveries concerning condensation polymerization reactions, such as those by Carruthers and Flory (cf. reference *45* for an historical narrative), and their synthetic procedure closely resembles descriptions of polyamide synthesis. But the conditions under which one performs the synthesis of polyamides via condensation will lead to decomposition of amino acids. For example, syntheses performed in the manner described by Harada and Fox (*27*), in which the dry amino acids are melted at 180°C and then allowed to react for several hours at 170°C, result in an enormous degree of decomposition: approximately 99% of the initial mass charged to the Schlenk tube is lost as volatiles (e.g. NH3, CO_2, H_2O), and one is left with a brittle resin that is insoluble in all solvents save formic acid or o-cresol. But if the glutamic acid melt is prepared and allowed to react at 160°C, there is very little decomposition, and one recovers solid material whose solubility behavior and response to spot test analyses are more consistent with those of peptides.

Much of the criticism directed at the Harada-Fox syntheses and thermal vent models in general arises from decomposition processes that occur once one exceeds a temperature of 160°C (*47*). The problem is particularly acute if one uses the syntheses of polyamides as a guide for setting reaction conditions, but it should be pointed out that Harada and Fox do mention in their papers that the reaction yields were improved at low temperature (*27,28*). Bulk samples of glutamic acid in our Schlenk tubes (25 mm diameter) take approximately 2 hours to melt at 160°C, but once this occurs the reaction proceeds without the discoloration of the melt (turning amber) described by Harada and Fox (*27,28*) or significant loss of mass not attributable to water being expelled. And finally, the polymerization of glutamic acid appears to proceed via a ring-opening mechanism: the work-up of melts that are quenched after reaction times of 2, 4, 8, 16 and 32 hrs reveal that glutamic acid is consumed after 4 hrs, followed by a steady increase of polymeric material up to 32 hrs.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

Templating in Synthetic Chemistry

In general, the outcome of chemical syntheses in homogeneous media is governed by the relative thermodynamic stability of all reaction outcomes, but the outcome of a given reaction can be directed towards a particular product by additives that act as a template during the reaction process (i.e. assembly of reactants into product material). The archetypical examples of these types of reactions are associated with macrocycle synthesis (*48*). For example, crown ethers are synthesized by linking two bifunctional linear molecules in a Williamson synthesis whose outcome is altered from the expected polyether (i.e. linear, with a product distribution representing a range of molecular size) to a small cyclic ether array of uniform size (Figure 2). This engineering of the reaction outcome is achieved by alkali metal cations in the reaction mixture, which coordinate the linear ether reactants and hold them in place during the coupling process. In other words, thermodynamic stability of products is superseded by a kinetic process, that is, the organization of molecular components on a template, during the reaction (*49*).

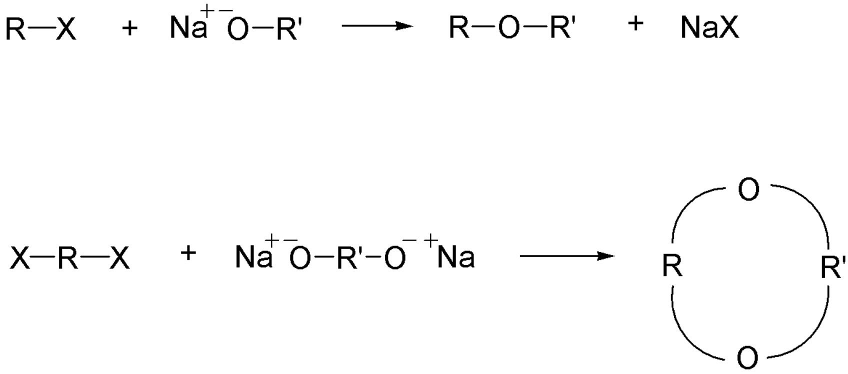


Figure 2. The Williamson Synthesis.The upper scheme illustrates the general method of coupling, whereas the bottom scheme depicts the route towards cyclic structures that are usually templated.

The synthesis of crown ethers represents an example of function following structure in the sense that the crown ethers owe their structure to reactions templated by a metal ion, which, in turn, results in their function as selective metal ion hosts. In many instances of synthetic biomimetic chemistry, the goal of the synthetic procedure is to replicate the selective binding of a substrate such as that found in enzyme active sites, and so the engineering of molecular recognition and/or selectivity by synthetic means is similarly achieved by using an ionic or molecular template during the assembly of the requisite synthons. A necessary condition in the designation of any given reaction as being templated is this topological control. And, in general, there is a kind of hierarchy among these host-guest combinations that may be used as a guide (*50*).

For example, a simple bimolecular complex might entail a dative interaction between two atomic or molecular species, such as the iodine-benzene charge-transfer complex, that features only minimal spatial organization. These complexes are weak and topologically indeterminate because the interacting pair

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

has a small number of functionalities participating in the interaction (mono-dentate, bi-dentate), and so spatial ordering is often observed only when these complexes are crystallized (e.g. 2:1 diselenane:iodine). Increasing the number of ligand atoms on the individual molecules allows the molecules to interact over more loci, which then imposes spatial constrains upon the resultant complex. Ethylenediamine tetraacetic acid (EDTA) is a familiar complexing agent whose affinity and binding stability towards cations is greater than other reagents such as ethylene diamine because the former is hexadentate and coordinates metal cations at six positions, as opposed to two in the case of ethylene diamine. Similarly, a podand is an open-chained polyether (and therefore an analogue of crown ethers) that, because of the multiple ligand atoms, can wrap itself around a cation guest. In other words, the type of biomimetic chemistry that occurs in this case is based upon the chelate effect.

With the formation of a polyether ring, as opposed to a linear array, it follows that crown ethers represent another degree of (spatial) order imposed upon a multi-dentate molecule. Additional spatial order may be imposed by capping crown ethers (e.g. lariat ethers, cryptands) or synthesizing spherical complexes (spherands). It is significant to note that with each step of increasing spatial order one observes a greater degree of selectivity of the so-called host towards a prospective guest (*48,50*). A catalogue of these various topological ligands is illustrated in Figure 3.

The jargon used in defining template-mediated reactions varies. One scheme of categorization relies on the reaction topology and defines "cyclization", "linear", or "interweaving" templates (*49*). The preceding description of crown ether synthesis is a representative example of cyclization templating, whereas the replication of DNA strands would correspond to a linear template reaction. In the interweaving process, the template acts as a kind of locus for two molecules that independently react and become entwined (no covalent bond between two independent molecules). The corresponding terms found in the literature of templated macrocycle synthesis (*48*) are "positive template", around which a synthon wraps prior to intramolecular coupling, and "negative template", which prevents intramolecular coupling by separating the reactive end groups of the synthon.

Coordination polymerization denotes a technique in which a monomer is directed towards the growing chain end through the formation of a coordination complex prior to linkage (*51*). For example, in the case of the so-called Ziegler-Natta catalysts that are used in the polymerization of dienes, the formation of a bimetallic complex in which the monomer is held in close proximity to the polymer's terminal end ensures that the polymer 'grows' in a linear fashion and that the monomer addition is subjected to stereochemical control (*51*). More selective techniques of polymer synthesis have been introduced since the discovery of the Ziegler catalysts, and these have attributes analogous to those found in host-guest complexes.

Templating, or 'molecular imprinting' (*52*), is currently used during the polymerization process in order to engineer selective sites in macromolecular adsorbants (*53*). The template functions in a manner identical to that described in the case of macrocycle synthesis, although the attributes of the template after synthesis in some cases differ. For example, it is mandatory that the template be

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

removable following the synthesis when the imprinted polymer network is to be used as a chromatography medium. But as in the case of the cyclization reactions that are used to synthesize host-guest pairs, the template interacts with a monomer via some functional group that is supplementary to the functional groups that participate in chain growth, and these newly coordinated monomers become part of the growing polymer network while fixed in a geometric order that is dictated by the template.

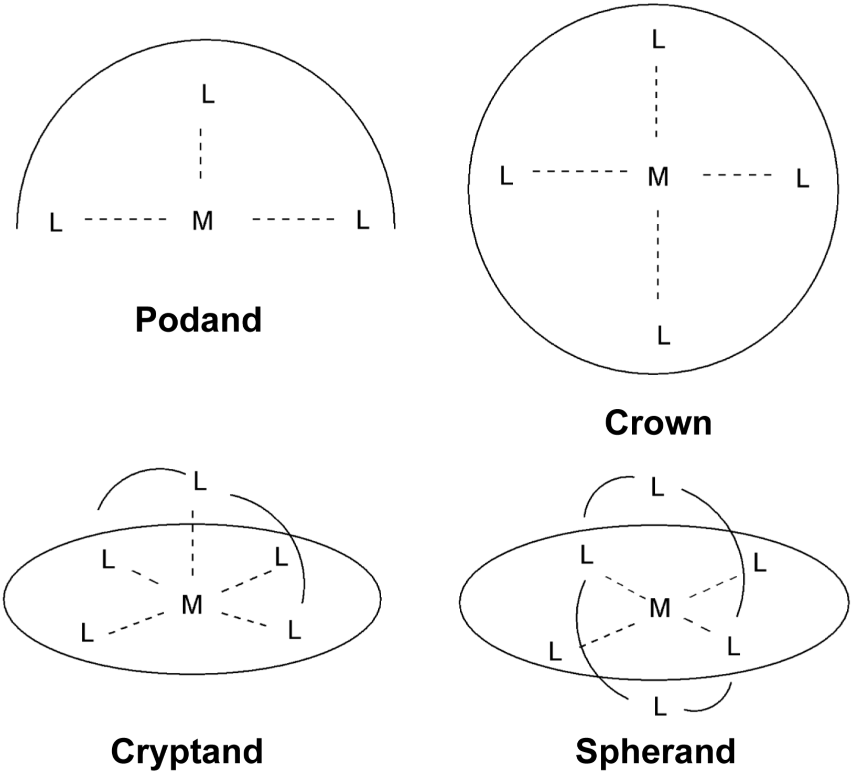


Figure 3. A catalogue of templated ligands of increasing level of encapsulation and specificity.

The imprinted adsorbants also seem to impart stability to their guest. Silica gels that are imprinted by and demonstrate specificity towards dye molecules also retard the degradation of the dye in the bound state, and similar experiments with a protein (cytochrome c) yielded an adsorbate that retained its function despite dry storage at room temperature, and harsh chemical treatment (*54*). This demonstrable resistance to degradation is relevant to abiotic origins of macromolecules and chemical evolution because it addresses some of the criticism raised concerning the stability of the organic precursors. Molecular imprinting may also be recognized as being at the root of one theory that attributes the chemical evolution of molecular information (*55*) to the mineral world: crystal structure of inorganic compounds imprints the organic milieu, which passes along the imprinting as the organic systems become more

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

complicated. One likewise sees the reflexive interplay between mineral and organic syntheses in the literature of biomineralization (*56,57*).

On the Prospect of Templating and the Chemical Evolution of Protein Motifs

Amino acids and peptides form well-defined coordination compounds with metal cations (*58,59*), and so it stands to reason that any polymerization of amino acids that occurs in an abiotic system will be governed by the geochemical contribution of salts to the reaction milieu. Evidence of such a templating effect has been presented in connection with the evaporative cycle experiments (*60*), which are also denoted salt-induced peptide formation reactions or the so-called seaside model of chemical evolution (*61*). Typical of these reactions is the mixing of amino acids in an aqueous solution containing $CuCl_2$ and NaCl (mole ratio of a single amino acid to Cu^{+2} and Na^+ is approximately 1:1:10), followed by evaporation of the water at 85-90°C and replacement of the water in 24 hr cycles. The products, which are dipeptides, are selected in the sense that certain pair sequences appear more often, and the preferences shown coincide with "three of the four most frequent homopeptide linkages in archaebacteria" (*62*).

The side chain group of amino acids provides a means to interact with a template while leaving the α- amino and carboxyl groups free to participate in chain propagation reactions, which is how molecular imprinting is achieved among the synthetic polymers (*51*). The metal binding groups of amino acids are diverse, and affinity of these ligands towards a metal ion is ultimately governed by their pK (*63*). In other words, the coordination of the amino acid by a metal ion template is determined by how much more strongly the metal binds to the amino acid's functional group than hydrogen ions, which in turn is determined by the solvent system in which the amino acids and metal ions are dissolved. And so it is plausible that co-polymerization of amino acids in the presence of metal ions would be imprinted as coordination complexes form, followed by ligand linkage *via* the insertion of peptide segments.

A molecular imprinting scenario may be illustrated using copper proteins as a model. The copper binding site that is found among the so-called 'blue' copper proteins features a motif in the form of a loop that contains three amino acids whose side chains function as ligands to the Cu^{+2} ion, $\text{Cys–X}_n\text{–His–X}_m\text{–Met}$, where X_i denotes a 'spacer' of i amino acid units. For example, n and m are 4 and 3, respectively, in the Type I copper protein azurin (*64*). One might therefore conceive of a primordial copper protein created abiotically as copper ions coordinated, for example, cysteine, histidine, and methionine, which were then linked to form a motif resembling the Type I binding site. The process, presented as such, does not connote efficiency, but a lack of efficiency is not necessarily inherent to molecular imprinting (*52*).

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

Condensation Polymerization in the Presence of Metal Ions

Modification of the Harada-Fox (*27*) Synthesis

In crystals obtained from aqueous solutions, glycine forms metal complexes as a bidentate ligand *via* its terminal carboxyl and amine functional groups, for example, $Zn(Gly)_2 \cdot H_2O$ (*65*). Non-chelate (unidentate) coordination by glycine does occur in two special cases (i.e. with Ag(I) and Pt(II); cf. reference *59*). Glycine and its oligomers also form stable complexes with metal ions in solution that, in turn, enhance the solubility of both members of the complex (*59*). And so by using the polymerization of glycine in a pyroglutamic acid melt as a benchmark, we can examine the prospect of templating by incorporating metal cations into the melt as the formate or acetate salt.

As a standard protocol replicating the Harada and Fox procedure, a glutamic acid and glycine mixture (molar ratio 2:1) is heated at 180°C for 1 hour, at which point the reaction tube containing the melt is transferred to a second block heater at the desired temperature (165-170°C) for the specified reaction time. An improved procedure, however, that greatly reduces pyrolytic decomposition of the amino acids entails using the 160°C temperature for both the melting process and subsequent reaction.

During trials intended to explore the role, if any, of the primordial atmosphere, we experimented with Schlenk reaction tubes outfitted with a Contat-Gockel trap (a saturated solution of $NaHCO_3$) that exchanges CO_2 and H_2O with the vapor over the melt. This method of poising the atmosphere above the melt prevented evaporative loss of water and resultant formation of a hardened resin, and from the standpoint of chemical evolution is consistent with an atmospheric model that is compositionally dominated by CO_2 and H_2O vapor. When performed under these conditions, the melt obtained during the polymerization of glycine in pyroglutamic acid typically remains a wheat color – it does not turn amber – and remains plastic. Work-up, following the procedures described by Harada and Fox (*27*), is thereby facilitated and yields the identical white polymeric precipitate.

The addition of metal ions to the melt affects the rate of polymerization and therefore the number average molecular weight of polyglycine produced in a Harada-Fox reaction scenario, and this result is consistent with the suggestion by Harada & Fox that the precise mechanism of polymerization in their mixture of pyroglutamic acid and glycine may be through ionic addition (*28*). But this same manipulation of rate and degree of polymerization can be achieved by varying the surface-to-volume ratio of the melt, and so the rate at which water can be physically removed from the melt *via* the interface is, at least in part, a factor. The metal ion does not seem to promote incorporation of glutamic acid into the chain, which is somewhat surprising given that metal ions are used as initiators for ring-opening polymerizations (in many cases using lactams as the monomer). Evidence for chain transfer linkage formation may be inferred from mass spectral data for peptides synthesized using glycyl-glycine as a starting material: the peptides produced do not grow by multiples of a two-glycine unit, as expected for a simple condensation process, but rather include odd-numbered oligomers.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

Complexation by the amino acids prior to polymerization alters the kinetics of chain propagation, which is consistent with the presumed catalytic role of the metal in these reactions (*51*), but amino acid salt formation has further implications from the standpoint of chemical evolution: the polymerizations may be performed at temperatures that more closely resemble those found on the Earth's surface. Amino acids form salts with carboxylic acids (e.g. formic and acetic), and these salts tend to melt at lower temperatures than the corresponding amino acid. For example, a mixture of the formate salts of glutamic acid and glycine (analogous to the first Harada-Fox melt, reference *27*) fuses at 150°C, and the yield of polymer after four hours exceeds the typical yield obtained with the corresponding melt of amino acids. This trend towards lower melting points continues as one increases the number of amino acids in the mixture: as a mixture of their formate salt, the Fox-Harada mixture of 16 amino acids (cf. reference *28*) is a viscous paste at room temperature. The formation of peptides and ultimately precipitation of macromolecules from a concentrated aqueous solution of amino acid salts can be effected at 50°C.

Resolution and Discrimination of Products in the Peptide Mixtures

From the standpoint of chemical evolution, the preceding results are relevant to reconciling abiotic amino acid polymerization reactions with temperatures attainable on the Earth's surface. But from the standpoint of biomimetic chemistry and the chemical evolution of functional proto-enzymes (i.e. motifs), they do not, and in fact can not, provide insight as to whether the metal ion is directing the reaction outcome towards a specific set of (imprinted) macromolecules. And so one has to implement a different type of experimental assay, namely two-dimensional mapping methods such as those employed in proteomic studies (*66*). The goal in proteomic research is to assess gene expression, namely the protein products, in an organism in a temporal context, and so a large number of diverse proteins must be assayed in a single sample: a situation analogous to that found in a synthetic chemical evolution experiment. A two-dimensional 'map' of the protein mixture is generated by chromatographic means with two or more solvents (*67*), by combining electrophoresis and chromatography, or, alternatively, isoelectric focusing and electrophoresis. Individual proteins will migrate according to their charge, size and chemical affinity towards the medium, and the resultant pattern of proteins reveals a compositional profile of the mixture. Once resolved, the proteins may be eluted for further analysis.

Figure 4 illustrates how two-dimensional mapping may be used to assess changes in the peptide composition of a melt following a Harada-Fox type of synthesis containing 16 amino acids (*28*). The maps were obtained on a cellulose substrate with the first dimension corresponding to electrophoresis at pH 3.5, followed by a second dimension chromatograph using n-butanol:pyridine:acetic acid:water as a solvent. The patterns that are obtained from this small sample of reactions are clearly different, which presumably reflects the effect of metal ion addition to the melt. Chromatographs 4a-4c correspond to variations in the melt after incubation for 4 hrs at 170°C, and there

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

is a marked change when Cu^{+2} is added to the melt. A fourth map, Figure 4d, represents the peptide profile of the same 16 amino acid mixture after incubation at 60°C in an ammonium formate solution for seven days.

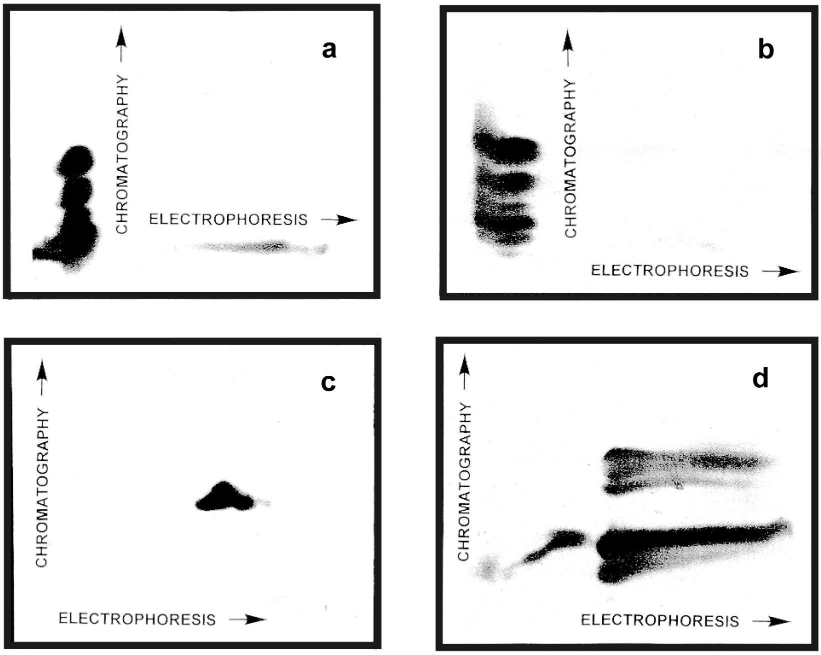


Figure 4. Two-dimensional maps of amino acid melts and their proteinoid products. First dimension: thin layer electrophoresis; second dimension: thin layer chromatography. (a) Peptide map obtained following the incubation of glycine in a glutamic acid melt at 170 ℃ for 4 hrs; (b) Peptide map obtained following the incubation of the same amino acid melt with Mg^{+2} *added as the acetate salt; (c) Peptide map obtained following the incubation of the same amino acid melt with* Cu^{+2} *added as the acetate salt; (d) Peptide map obtained following a replication of the Harada-Fox experiment (28) using the 16 amino acid mixture, but at far lower temperature: 60 ℃.*

Conclusion

This literature survey has demonstrated that there exist parallels between the literatures of synthetic chemistry and chemical evolution and that they are, in fact, complementary. One can, for example, identify individual techniques of laboratory polymer synthesis in each of the somewhat broader and molecularly more encompassing models of chemical evolution. The variety of experimental tests applied to models of chemical evolution, both with respect to small molecules and macromolecules, plus the successful synthesis of biomimetic molecular devices by means and reagents that were not likely to have existed on the primordial Earth, suggest that no one process may be attributed to the origin

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

of life, at least with respect to its requisite molecules. In other words, chemical evolution in all likelihood resembles hereditary evolution in which statistics (and the skewing thereof by exogenous factors) determines the compositional profile of the system – much as one would describe the pooled reactions of metabolism.

What does seem to recur and therefore stand out as an important factor, possibly as a modifier of the statistics, is molecular imprinting. Molecular imprinting affects a synthetic process *via* kinetic factors (as opposed to thermodynamic) and therefore does alter the statistics of the process, in effect, directing outcomes. The discussion in previous sections concerned molecular imprinting in polymers as a means towards the creation of specific binding sites that are suggestive of proto-enzymes, but molecular imprinting is likewise implicated at the onset of chemical evolution, that is, during cosmochemical events, because certain small molecules (e.g. polyaromatic hydrocarbons) seem to be closely associated with a particular type of interstellar dust (cf. reference *16*). It stands to reason therefore that arguments to the effect that life emerged from combinatorial events (*68*), as opposed to a direct route, are valid, and such arguments applied to so-called chemical evolution thereby merge seamlessly both with the chemistry of metabolic processes and with theories of organismal evolution, such as that described by Gould (*69,70*).

References

1. *Colloid Chemistry;* Alexander, J., Ed.; The Chemical Catalog Company: New York, 1929; Volume 2, p 327.
2. Drexler, K.E. *Engines of Creation;* Doubleday: Garden City, NJ, 1986.
3. Tanford, C.; Reynolds, J. *Nature's Robots. A History of Proteins;* Oxford University Press: New York, 2001.
4. Breslow, R. *Chem. Soc. Rev.* **1972,** *1,* 553–580.
5. Breslow, R. *Acc. Chem. Res.* **1980,** *13,* 170–177.
6. *Biomimetic Chemisty;* Dolphin, D.; McKenna, C.; Murakami, Y.; Tabushi, I., Eds.; American Chemical Society: Washington, DC, 1980.
7. Calvin, M. *Chemical Evolution;* Oxford University Press: New York, 1969.
8. Lemmon, R. M. *Chem. Rev.* **1970,** *70,* 95–109.
9. *Encyclopedia of Supramolecular Chemistry;* Atwood, J. L.; Steed, J. W., Eds.; Marcel Dekker, Inc.: New York, 2004; Volume 2, p 1493.
10. *The RNA World;* Gesteland, R. F.; Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Plainview, NY, 1993.
11. Woese, C. *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 6854–6859.
12. Corey, E. J. *Pure Appl. Chem.* **1967,** *14,* 19–38.
13. Fuhrhop, J.; Penzlin, G. *Organic Synthesis. Concepts, Methods, Starting Materials;* Verlag Chemie GmbH: Weinheim, Germany, 1986.
14. Corey, E. J.; Cheng, X.-M. *The Logic of Chemical Synthesis;* John Wiley & Sons: New York, NY, 1989.
15. Jaschek, C.; Jaschek, M. *The Behavior of Chemical Elements in Stars;* Cambridge University Press: Cambridge, UK, 1995.
16. *Dust and Chemistry in Astronomy;* Millar, T. J.; Williams, D. A., Eds.; Institute of Physics Publishing: Bristol, UK, 1993.

17. *Planet Formation. Theory, Observations, and Experiments;* Klahr, H.; Brandner, W., Eds.; Cambridge University Press: Cambridge, UK, 2006.
18. *Comets and the Origin and Evolution of Life, 2^{nd} Ed.;* Thomas, P. J.; Hicks, R. D.; Chyba, C. F.; McKay, C. P., Eds.; Springer-Verlag: Berlin, Germany, 2006.
19. *The Origin and Evolution of Planetary Atmospheres;* Henderson-Sellers, A., Ed.; Adam Hilger Ltd.: Bristol, UK, 1983.
20. Budyko, M. I.; Ronov, A. B.; Yanshin, A. L. *History of the Earth's Atmosphere;* Springer-Verlag: Berlin, Germany, 1985.
21. Rollinson, H. *Early Earth Systems. A Geochemical Approach;* Blackwell: Malden, MA, 2007.
22. Miller, S. L. *J. Am. Chem. Soc.* **1955,** *77,* 2351–2361.
23. Miller, S. L.; Urey, H. C. *Science* **1959,** *130,* 245–251.
24. Schäfer, G.; Engelhard, M.; Müller, V. *Microbiology and Molecular Biology Reviews* **1999,** *63,* 570.
25. Oparin, A. I. *The Origin of Life on Earth;* Oliver & Boyd: London, UK, 1957.
26. Haldane, J. B. *Rationalist Ann.* **1929,** *148,* 3–10.
27. Harada, K.; Fox, S. W. *J. Am. Chem. Soc.* **1958,** *80,* 2694–2697.
28. Fox, S. W.; Harada, K. *J. Am. Chem. Soc.* **1960,** *82,* 3745–3751.
29. Woese, C. R. In *The Origins of Life and Evolution;* Halvorson, H. O.; van Holde, K. E., Eds.; Alan R. Liss, Inc.: New York, 1980; pp 65–76.
30. O'Dowd, C. D.; de Leeuk, G. *Phil. Trans. R. Soc. A* **2007,** *365,* 1753–1774.
31. Dobson, C. M.; Ellison, G. B.; Tuck, A. F.; Vaida, V. *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 11864–11868.
32. Tuck, A. *Surveys Geophys.* **2002,** *23,* 379–409.
33. Baross, J. A.; Hoffman, S. E. *Origins of Life* **1985,** *15,* 327–345.
34. Daniel, I.; Oger, P.; Winter, R. *Chem. Soc. Rev.* **2006,** *35,* 858–875.
35. Miller, S. L.; Bada, J. L. *Nature* **1988,** *334,* 609–611.
36. *Polymerization and Polycondensation Processes;* Platzer, N. A., Ed.; American Chemical Society: Washington, DC, 1962.
37. *Chemical Synthesis Using Supercritical Fluids;* Jessop, P. G.; Leitner, W., Eds.; Wiley-VCH: Weinheim, Germany, 1999; p 297.
38. Holm, N. G.; Charlou, J. L. *Earth Planet. Sci. Lett.* **2001,** *191,* 1–8.
39. Lindsay, J. F.; Brasier, M. D.; McLoughlin, N.; Green, O. R.; Fogel, M.; Steele, A.; Mertzman, S. A. *Precambrian Research* **2005,** *143,* 1–22.
40. Burke, A. L.; Maier, G.; DeSimone, J. M. *Polym. Mater. Sci. Eng.* **1996,** *74,* 248–249.
41. Shock, E. I. *Geochim. Cosmochim. Acta* **1992,** *56,* 3481–3491.
42. Shock, E. I. *Geochim. Cosmochim. Acta* **1993,** *57,* 3341–3349.
43. Klinger, D.; Berg, J.; Vogel, H. *J. Supercrit. Fluids* **2007,** *43,* 112–119.
44. Pearce, E. M.; Wright, C. E.; Bordoloi, B. K. *Laboratory Experiments in Polymer Synthesis and Characterization;* The Pennsylvania State University: University Park, PA, 1979.
45. Flory, P.J. *Principles of Polymer Chemistry;* Cornell University Press: Ithaca, NY, 1953; p 94.
46. Staudinger, H.; Klever, W.; Kober, P. *Liebigs Ann.* **1910,** *374,* 1–39.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

47. Wills, C.; Bada, J. *The Spark of Life. Darwin and the Primeval Soup;* Perseus Publishing: Cambridge, MA, 2000; p 54ff.
48. Gerbeleu, N.V.; Arion, V.B.; Burgess, J. *Template Synthesis of Macrocyclic Compounds;* Wiley-VCH: Weinheim, Germany, 1999.
49. Diederich, F.; Stang, P.J. *Templated Organic Synthesis;* Wiley-VCH: Weinheim, Germany, 2000.
50. Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry;* John Wiley & Sons: Chichester, UK, 2000.
51. Billmeyer, F.W. *Textbook of Polymer Science, 2nd. ed.;* John Wiley & Sons: New York, 1971; p 319.
52. Steinke, J.; Sherrington, D. C.; Dunkin, I. R. *Adv. Polymer Sci.* **1995,** *123,* 81–125.
53. Striegler, S. *J. Chrom. B* **2004,** *804,* 183–195.
54. Dickey, F. H. *J. Phys. Chem.* **1955,** *59,* 695–707.
55. Cairns-Smith, A. G. *Genetic Takeover and the Mineral Origins of Life;* Cambridge University Press: Cambridge, UK, 1982.
56. *Biomimetic Materials Chemistry;* Mann, S., Ed.; Wiley-VCH: New York, 1996.
57. Whilton, N. T.; Vickers, P. J.; Mann, S. *J. Mater. Chem.* **1997,** *7,* 1623–1630.
58. Greenstein, J. P.; Winitz, M. *Chemistry of the Amino Acids;* John Wiley & Sons: New York, 1961; Volume 1, p 569.
59. *Inorganic Biochemistry;* Eichhorn, G. L., Ed.; Elsevier Scientific: Amsterdam, Netherlands, 1973; Volume 1, p 162.
60. Saetia, S.; Liedl, K. R.; Eder, A. H.; Rode, B. M. *Origins Life Evol. Biosphere* **1993,** *23,* 167–176.
61. *Geochemistry and the Origin of Life;* Nakashima, S.; Maruyama, S.; Brack, A.; Windley, B. F., Eds.; Universal Academy Press: Tokyo, Japan, 2001.
62. Rode, B. M.; Eder, A. H.; Yongyai, Y. *Inorg. Chim. Acta* **1997,** *254,* 309–314.
63. Glusker, J. P. *Adv. Protein Chem.* **1991,** *42,* 1–76.
64. Baker, E. N. *J. Mol. Biol.* **1988,** *203,* 1071–1095.
65. Low, B. W.; Hirshfeld, F. L.; Richards, F. M. *J. Am. Chem. Soc.* **1959,** *81,* 4412–4416.
66. *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods;* Rabilloud, T., Ed.; Springer Verlag: Berlin, Germany, 2000.
67. von Arx, E.; Neher, R. *J. Chromatog.* **1963,** *12,* 329–341.
68. Dose, K. *Adv. Space Res.* **1986,** *6,* 181–186.
69. Gould, S. J. *The Structure of Evolutionary Theory;* J.; Harvard University Press: Cambridge, MA, 2002; Part 2.
70. Gould, S. J. *Punctuated Equilibrium;* Harvard Univeristy Press: Cambridge, MA, 2007.

Downloaded by UNIV OF DELAWARE MORRIS LIB on June 28, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch016

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

Chapter 17

Better Evolution Through Chemistry: Rapid Evolution Driven by Human Changes to the Chemical Environment

Stephen R. Palumbi

Department of Biology, Stanford University, Stanford, CA 94305

Evolution occurs when phenotypic variation among organisms causes differences in reproductive success, and when these phenotype differences are passed on to offspring. Human mediated changes in local and global environments can dramatically change levels of natural selection and can lead to rapid evolutionary change. Examples of pesticide and antibiotic resistance have grown along with the use of these chemicals in agriculture in medicine, especially when evolutionary arms races occur. Though small organisms with fast generation times are most likely to provide clear examples of rapid evolution, there are examples among larger-bodied species including humans. A pervasive chemical modification of the environment due to increased CO_2 levels is threatening a wide array of plant and animal life. Though some species are already evolving in the face of such changes, climate shifts due to CO_2 increases are occurring ten times faster than during a typical glacial cycle. Species with long generation times, or complex morphological traits are least likely to evolve fast enough to keep pace with this rapid evolutionary challenge.

In Chemical Evolution II: From the Origins of Life to Modern Society; Zaikowski, L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2009.

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

Charles Darwin laid out the basic rules by which populations evolve, emphasizing that three major elements are required: variation among individuals, a link between variation and reproductive fitness, and heritability of the variation across generations (see *1* for review). Darwin traveled the world to accumulate evidence for the evolution of populations by natural selection, and found it in fossils and populations separated for millions of years (*2*). But 150 years later, we can often see evolution occurring on our very doorsteps and can document rates of evolutionary change that would have astounded Darwin.

Some of these examples of rapid evolutionary change derive from the strong impact humans have on the environment. By changing environments on local or global scales, human activity exerts new selective forces, or strengthens forces that might have been already acting (*1,3*). Study of the evolutionary responses to human environmental change have provided key examples of adaptive evolution. In many cases, rapid adaptation is seen among pest or disease species, and sparks a complex response from human industry that often results in an evolutionary arms race. Increasing attention is focused on the ecological effects of global climate change - with a major question being whether species are likely to evolve fast enough to keep up with current environmental shifts.

Down on the Farm

Agricultural uses of pesticides have increased tremendously from the beginnings of this industry in the 1950's. Annual expenditures on pesticides nearly doubled in the United States from 1980 to 2000, with increases in herbicide, insecticide and fungicide use (Figure 1; *4*). At the same time, the number of insect species that have evolved resistance to at least one insecticide has reached about 500 by the year 2000, with some insects evolving powerful resistance to multiple, commonly used chemicals (*5,6*).

Study of the molecular mechanisms of resistance shows some cases of evolutionary tinkering with normal cellular machinery (*7*). For example, resistance to the BT toxin Cry1Ac, genetically engineered into cotton, derives from evolution of a cadherin receptor gene so that it no longer binds to the toxin molecule (*8,9*). Other genetic changes include the upregulation of genes that detoxify pesticides, or the duplication of these genes so that they can produce more product. However, the availability of genome sequences from some insects also makes it possible to discover more radical adaptive evolution.

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

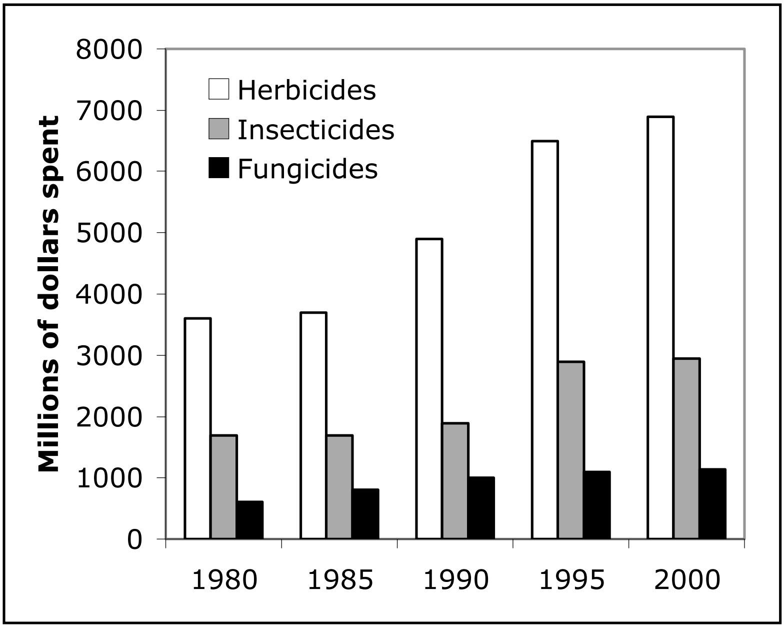


Figure 1: U.S. agricultural expeditures on pesticides have grown steadily. Data are from reference 4.

In the fruit fly *Drosophila melanogaster* the mobile genetic element Doc1420 has recently jumped from gene to gene. When such an element moves in the genome and inserts itself inside a different gene, it often disrupts that gene's transcription. In one case, however, Doc1420 inserted itself inside the second exon of a particular fly gene, creating a new gene that contained the original first exon and a modifed second exon made up by half the original second exon and half by the mobile element itself (Figure 2). Over the past few decades, the frequency of this altered gene in the *D. melanogaster* population has increased remarkably, suggesting that it provides some selective advantage. Further work has shown that the new gene confers strong resistance to organophosphate insecticides (*10*). In this case, a radical genetic change appears to be adaptive in the presence of a modified chemical environment.

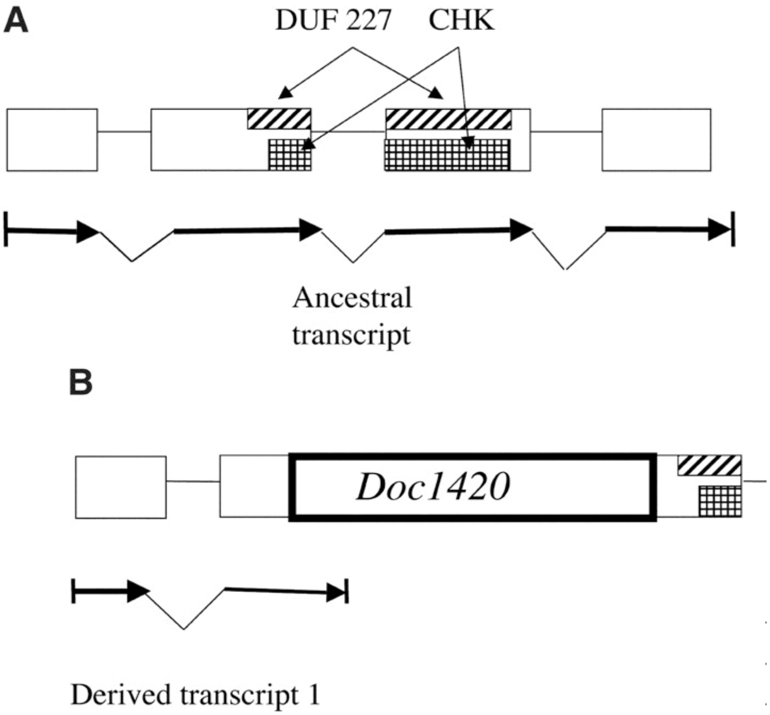


Figure 2: A gene in Drosophila (a) has had a transposable element Doc1420 inserted into the second exon (b), generating a new gene that confers organophosphate resistance, is evolving rapidly, and has increased in abundance in the past few decades (Reproduced with permission from reference 10. Copyright 2005 AAAS.)

Public Heath and Evolution

Insecticides have often been used for public health purposes, and have resulted in both widespread chemical usage and widespread evolution. The first use of DDT to control an arthropod infestation came in Naples in 1944. It was used by the invading Allied army to stop an infestation of lice that transmitted typhus (*3*). Such was the success of this program, that DDT was subsequently used all over the world to control insect vectors of disease. Malaria was just about eradicated in the United States in the 1950s by a combination of DDT and wetland drainage. However, attempts to repeat this success on the global scale largely failed because of the rapid evolution of DDT resistance by mosquitoes. In addition, the single celled blood parasite that causes malaria in humans (*Plasmodium falciparum*) has also evolved resistance to a battery of drugs used to treat it (see the readable review in *11*).

Today, malaria remains a critical global disease, with over 1 million deaths a year attributed to drug resistant *Plasmodium* (see *12*). Recently a new strategy

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

has been deployed that has had some success. DDT is again being used to combat mosquitoes, but now the treatment does not involve wholesale spraying, but instead focuses on mosquito netting impregnated with DDT. This approach provides local protection against mosquitoes without exposing the whole mosquito population to such high DDT levels that DDT resistance quickly evolves. In addition, a new natural product based drug, Artemisin, appears to be effective in combating even multi-drug resistant *Plasmodium*. In Rwanda, treatment with DDT nets and Artemisin has helped drop hospital cases of malaria from 10,000 a year to under 4,000 cases a year (*13*). However, signs of evolution are already occurring and partial resistance to Artemisin has been recorded (*14*).

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

Evolutionary Arms Races

Evolutionary co-adaptation of predator and prey is a common occurrence in the history of life. This can lead to evolutionary arms races in which the evolution of better predation ability exerts selective pressure on prey for better escape, which exerts selection on the predator for higher abilities, etc. Similar evolutionary arms races have played out between disease microorganisms and human antibiotic use. In these cases, powerful antibiotics are invented and used widely, leading to selection on bacterial disease populations to adapt to them. The failure of antibiotics then leads to a search for a new antibiotic, which, when used widely, generates new adaptation, etc.

A long series of adaptive evolutionary changes and antibiotic shifts has occurred during treatment of *Staphylococcus aureas* infections (Table I). At first, penicillin was nearly completely effective against these infections. But rising resistance rates in the 1950s caused a wholesale switch by hospitals to using a second drug, methicillin. Methicillin resistance was first reported in Cairo in 1961, and has been increasing steadily in the last few decades. People with a Methicillin Resistant Staphylococcus Aureus (MRSA) infection were treated with vancomycin instead. By the 2000s, MRSA became so common that methicillin is now rarely used. Vancomycin resistance in *S. aureus* was reported in Japan in 1996, and in 2000 a new drug, Linezolid, was approved for use against vancomycin resistant MRSA. Linezolid resistance was reported in 2002, and incidence of resistant infections, while usually under 1% by 2008, appears to be growing.

Table I: History of Antibiotic Treatment of *Staphylococcus aureus* Infections.

Year	*Antibiotic history*
1943	Penicillin widely available
1947	First resistance reported
1960s	Switch to Methicillin
1961	Methicillin resistance reported
1980s	Methicillin resistance increasing, use of Vancomycin begins
1990s	Methicillin resistance up to 35%, Vancomycin use increasing
1996	Vancomycin resistance reported
2000	Linezolid approved
2002	Linezolid resistance reported
2006	Linezolid resistance rates 0.1% - 4% in hospitals

These shifts in drugs mean that *S. aureus* infections are currently still treatable, but just as in other evolutionary arms races, the cycle of antibiotic -> resistance evolution -> new antibiotic generates escalation. For bacteria, the escalation is in the strength of antibiotic resistance. For the human community, escalation often means increases in cost of health care. Partially these increases come from increased hospital stays when an infection is resistant to treatment. Other costs, however, include the drugs themselves, which are expensive to develop and produce. Though Table II is by no means a broad survey of prescription costs, one internet supplier of pharmaceuticals in 2008 lists the per-dose price of drugs used to combat *S. aureus* as a low of $0.69 for non-resistant infections, up to $42.75 for Linezolid. This nearly two-order of magnitude increase is part of the human price for antibiotic resistance.

Other costs are also high. MRSA is a serious disease with a high mortality rate, and aggressive treatment is required. Kim et al. (*15*) list the average cost of treating MRSA infections contracted outside of hospitals at about $14,700 per case (*16*). The 100,000 out-of-hospital cases of MRSA cost about $1.4 billion. Overall, treating antibiotic resistance of *S. aureus* has been estimated to cost $24-$30 billion a year in the United States (*3*).

Table II: Survey of Prescription Costs per Dose from Internet Sources

Name	*cost/#doses*	*cost per dose*
Amoxicillin	$69.00 / 100	$0.69
Methicillin	off the market	N/A
Vancomycin	$186.50 / 28	$6.64
Linezolid	$855.00 / 20	$42.75

SOURCE: Prices from RxPOP.com April, 2008.

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

Evolution Writ Large

The forgoing examples have some strong similarities. In most cases, the species involved are small-bodied with short generation times and large population sizes. These traits tend to enhance the ability of a population to respond to strong natural selection quickly. However, evidence of evolution of larger species also exists, even in humans.

Among traits that may be considered adaptive in humans, one that received considerable study is the ability to digest lactose, the major sugar contained in milk. Most people express the major gene during infancy. However, after weaning, the gene (called lactase) is largely shut off in most humans. An exception occurs in populations of northern European origin, in which a single substitution in the regulatory region of the lactase gene is thought to retain the expression of this protein into adulthood, and to confer the ability to comfortably digest a diet high in dairy products. These populations also show a strong archeological record of cattle husbandry, and a strong element of dairy products in adult diets. A map of the prevalence of the allele that retains lactase production is very similar to the map of prevalence of cattle raising in Europe (Figure 3), suggesting that a change in diet sparked the evolution of this metabolic change, and that evolution by natural selection in the human population over the past 10,000 years is responsible for the ability of northern European adults to digest lactose (*17*).

However, another possibility exists: that lactase expression changed for other reasons, just by random genetic drift, for example. If this were true, then a high dairy diet might have been permitted by a prior genetic change. These two possibilities boil down to asking which came first, widespread cattle farming or the allele for adult lactase expression? Recently, this question was directly answered in a study of the DNA of bones from northern European populations of 5,000 - 7,000 years ago. Though there were only 8 alleles of the lactase gene characterized in this study from this period, only one had the substitution that shows it to confer adult expression of lactase (*18*). Thus the gene was present but not yet prevalent when cattle raising was first spreading widely in northern Europe. This evidence suggests that diet shifted before genetic evolution occurred, and that natural selection for adult digestion of lactose was responsible for the high frequency of this allele.

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

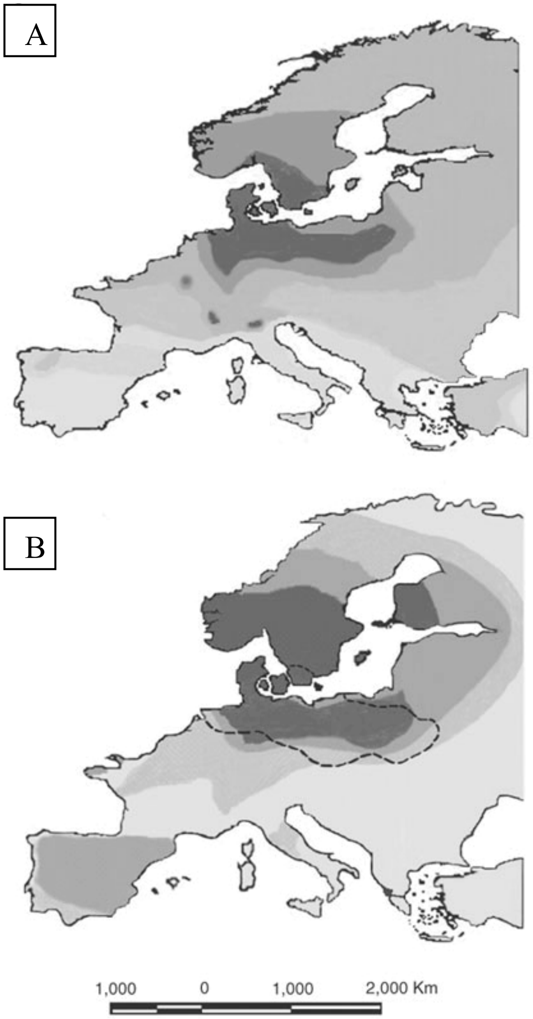


Figure 3: The distribution (dark shading) of prehistoric cattle ranching (a) and the genetic variant of the lactase gene promoter that allows it to be expressed in adults (b).The dashed line in b shows the region of intense cattle ranching activity from a. (Adapted with permission from reference 17. Copyright 2003 Nature Publishing Group.)

Climate Change and the Evolutionary Marathon

One of the most prevalent human-made chemicals is CO_2, produced so prodigiously that levels of CO_2 in the atmosphere and the oceans are far higher than they would be naturally. The resulting increases in global temperature, caused by greenhouse heating of the atmosphere, have begun to change overall temperature levels, the extent of arctic sea ice and sea level (*19*). Though these changes in climate have just begun, populations are beginning to evolve as a result (*19,20*). One example, involves the shift in wintering grounds of the European blackcap, a small bird that breeds in southern Germany and typically winters in Spain. Recently, warming of Great Britain has allowed a population of blackcaps to successfully overwinter there, a behavioral shift to climate change (*21*). Because individuals that overwinter in England arrive earlier on the breeding ground than those that overwinter in Spain, members of the English subpopulation tend to breed assortatively with one another, and produce more offspring. Over time, such a reproductive advantage may increase the English group relative to the Spanish group or even generate a new species (*22*). A second example comes from the Canadian Yukon, where temperatures have been increasing by an average of 2° C per decade. In this environment, as Spring weather has come more quickly, red squirrels have been breeding sooner. An analysis of phenotypic variance suggests that much of this shift is due to acclimation of individuals in response to temperature. But a substantial fraction (about 13%) of the response is genetically based, showing evolution for earlier breeding in this population (*23*).

These examples will probably be followed by others in which climate changes induce evolutionary shifts in populations. However, of concern is the rate at which evolution due to climatic changes can occur, and whether global warming, ocean acidification and sea level rise will outstrip the ability of populations to evolve.

The environment of the recent 2.5 million years has been replete with climate changes as glacial cycles have come and gone. One possibility is that current animal, plant and other species have already been subjected to many rounds of global climate change, and that the ones that persist now are the ones that can evolve quickly enough to persist. However, current climate changes are far faster than glacial cycles. For example, global temperatures increased about 7° C over the past 10,000 years since that last glaciation, an average rate of about 0.07° C per century (rates in individual habitats will vary). By contrast, global temperature has increased about 0.8° C in the past 110 years, ten times faster than glacial rates.

So far, there is no general theory that can estimate the maximum rate of evolution that can be sustained by a population, and so there is no way to predict which species may be left behind in the upcoming evolutionary marathon. However, examples of rapid evolutionary change, and basic evolutionary principles, suggest the kinds of species or traits that are likely to evolve most quickly.

Species with short generation times, large population sizes and small body size show the fastest rates of evolution in human-altered environments. Disease organisms, weeds and pests show high ability to adapt. Larger species with smaller populations and longer generation times can also evolve, but here the best examples come from behavioral shifts (such as migration pathways) or shifts in simple traits such as lactose digestion. Complex morphological or life history traits that depend on many genes are less likely to evolve quickly in small-population, long-generation species. These are the species, including most of the large animals and plants, that are most at risk of extinction due to poor ability to adapt evolutionarily to global change.

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

The implications of these simple evolutionary rules are not particularly surprising: the species or traits that can evolve the most quickly are most likely to be able to change during rapid climate shifts. The species or traits that can evolve the most quickly form a list which is also not that surprising - evolutionary biologists have long known the basic rules by which evolutionary rates are set (*24*). The same theoretical framework for evolution also suggests some caveats to the above: species or traits with very low genetic variation will also be slow to evolve, for example, no matter what their generation time or population size. Nevertheless, the same traits that tend to place species at risk ecologically also tend to result in slow evolution. The combination of these ecological and evolutionary constraints creates a list of species at risk ecologically that is very similar to the list that is at risk evolutionarily: large bodied species with long generation times and small populations. Species such as these - for example, large mammals, birds and reptiles - are not the only ones at ecological risk due to climate change, but they are the ones also least likely to be able to mount an evolutionary response just at the time when one is needed most.

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

References

1. Palumbi, S. R. Evolution - Humans as the world's greatest evolutionary force. *Science* **2001,** *293,* 1786–1790.
2. Darwin, C. R. *On the Origin of Species by Natural Selection;* D. Appleton: New York, 1872.
3. Palumbi, S. R. *The Evolution Explosion: How Humans Cause Rapid Evolutionary Change;* W. W. Norton: New York, 2001.
4. Annual User Expenditures on Pesticides in the U.S. by Pesticide Type, 1982 - 2001 Estimates, All Market Sectors US EPA 2001.
5. Groeters, F. R. Insecticide resistance. *Trends Ecol. Evol.* **1995,** *10,* 164.
6. McKenzie, J. A. *Ecological and evolutionary aspects of insecticide reistance. Environmental Intelligence Unit R.G.;* Landes/Academic Press: Austin, TX, 1996.
7. McKenzie, J. A.; Batterham, P. The genetic, molecular and phenotypic consequences of selection for insecticide resistance. *Trends Ecol. Evol.* **1994,** *9,* 166–169.
8. Tabashnik, B. E. Evolution of resistance to Bacillus thuringiensis. *Annu. Rev. Entomol.* **1994,** *39,* 47–79.
9. Tabashnik, B. E.; Liu, Y. –B.; Finson, N.; Masson, N.; Heckel, D. G. One gene in diamondback moth confers resistance to four Bacillus thuringiensis toxins. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94,* 1640–1644.
10. Aminetzach, T.; Macpherson, J.; Petrov, D. Pesticide resistance via transposon-mediated adaptive gene truncation in Drosophila. *Science* **2005,** *309,* 764–767.
11. Desowitz, R. S. *The Malaria Capers: More tales of parasites and people, research and reality;* W.W. Norton: New York, 1991.
12. URL: http://www.map.ox.ac.uk .

13. Jambou, R.; Legrand, E.; Niang, M.; Khim, N.; Lim, P.; Volney, B.; Ekala, M.; Bouchier, C.; Esterre, P.; Fandaur, T.; Mercereau-Puilalon, O. Resistance of Plasmodium falciparum field isolates ti in-vitro artemether and point mutations of the SERCA-type PfATPase6. *Lancet* **2005,** *366,* 1960–1963.
14. Afonso, A.; Hunt, P.; Cheesman, S.; Alves, A. C.; Cunha, C. V.; Rosário, Vd.; Cravo, P. Malaria Parasites Can Develop Stable Resistance to Artemisinin but Lack Mutations in Candidate Genes atp6 (Encoding the Sarcoplasmic and Endoplasmic Reticulum Ca2+ ATPase), tctp, mdr1, and cg10. *Antimicrob. Agents Chemother.* **2006,** *50,* 480–489.
15. Kim, T.; Oh, P.; Simor, A. The economic impact of methicillin-resistant Staphylociccus aureus in Canadian hospitals. *Infect. Control Hosp. Epidemiol.* **2001,** *22,* 99–104.
16. Gould, I. M. Costs of hospital-acquired methicillin-resistant Staphylococcus aureus (MRSA) and its control. *Int. J. Antomicrob. Agents* **2006,** *28,* 379–384.
17. Beja-Pereira, A.; Luikart, G.; England, P. R.; Bradley, D. G.; Jann, O. C.; Bertorelle, G.; Chamberlain, A. T.; Nunes, T. P.; Metodiev, S.; Ferrand, N.; Erhardt, G. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nature Genetics* **2003,** *35,* 311–313.
18. Burger, J.; Kirchner, M.; Bramanti, B.; Haak, W.; Thomas, M. G. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104,* 3736–3741.
19. IPCC 2007.
20. Bradshaw, W. E.; Holzapfe, C. M. Evolutionary Response to Rapid Climate Change. *Science* **2006,** *312,* 1477–1478.
21. Berthold, P.; Helbig, A. J.; Mohr, G.; Querner, U. Rapid microevolution of migratory behaviour in a wild bird species. *Nature* **1992,** *360,* 668–670.
22. Bearhop, S.; Fiedler, W.; Furness, R. W.; Votier, S. C.; Waldron, S.; Newton, J.; Bowen, G. J.; Berthold, P.; Farnsworth, K. Assortative Mating as a Mechanism for Rapid Evolution of a Migratory Divide. *Science* **2005,** *310,* 502–504.
23. Réale, D.; Berteaux, D.; McAdam, A.; Boutin, S. Lifetime selection on heritable life-history traits in a natural population of red squirrels. *Evolution* **2003,** *57,* 2416–2423.
24. Endler, J. A. *Natural Selection in the Wild;* Princeton University Press: Princeton, NJ, 1986.

Downloaded by UNIV OF GUELPH LIBRARY on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ch017

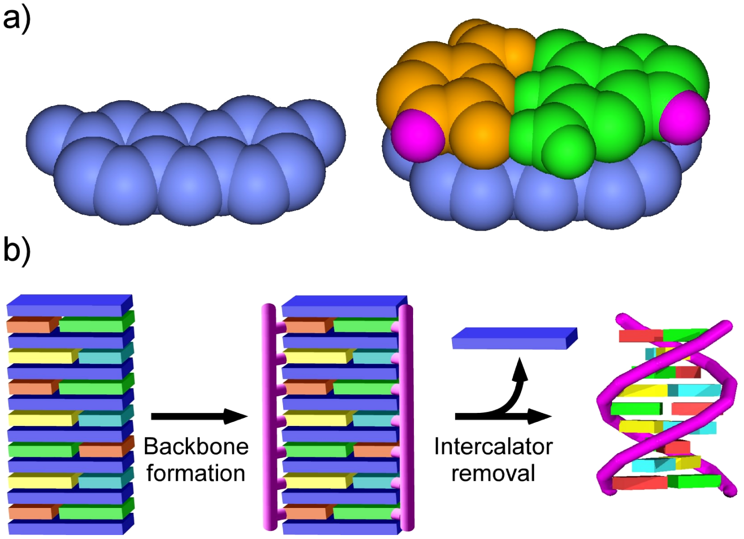


Plate 6.1. Schematic representation of the 'molecular midwife' hypothesis (57). a) A midwife molecule, proflavine (in blue) acts as a template for the formation of a Watson–Crick base pair. H-atoms of the nucleoside bases that will be replaced by C-atoms upon backbone formation are shown in magenta. b) In the case of molecular midwives with a greater association constant for base assembly than for self association, columnar stacks containing alternating midwife molecules and base assemblies spontaneously form under certain conditions. These columnar stacks preorganize the bases such that the introduction of a linkage chemistry leads to the formation of RNA-like polymers with a backbone length of 6.8 Å. Because the midwife molecules are only associated with the resulting polymers through non-covalent interactions, changes in solution conditions can lead to the removal of the intercalating midwives. (Reproduced with permission from reference 41. Copyright 2007 Verlag Helvetica Chimica Acta.)

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ap001

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ap001

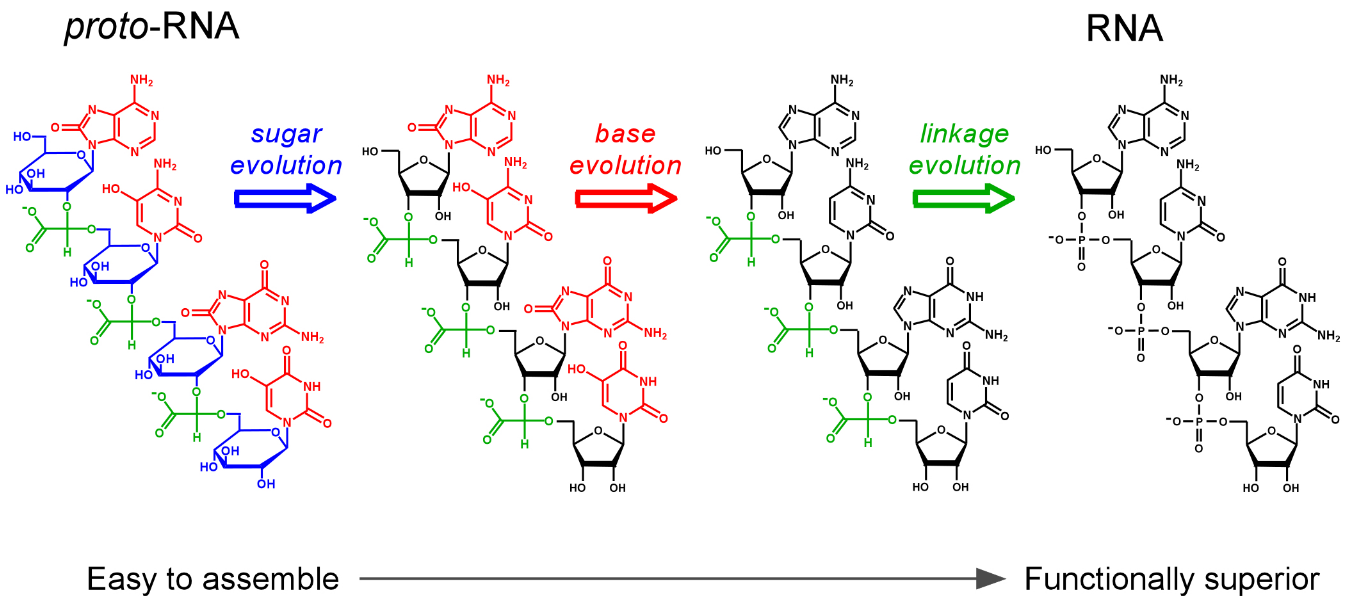

proto-RNA
RNA
sugar evolution
base evolution
linkage evolution
Easy to assemble
Functionally superior

Plate 6.2. The earliest proto-RNA polymers would have been composed of monomer units that were easy to assemble. Although the proto-RNA depicted here has a backbone comprised of a single type of sugar (blue) and a single linking molecule (green), in the absence of RNA or protein enzymes the earliest proto-RNAs were likely more heterogeneous. As proto-RNA sequences that promoted RNA synthesis were spontaneously generated and selected through Darwinian evolutionary processes, the populations of proto-RNAs would have become more homogeneous, stable, and functional, eventually leading to the catalyzed synthesis of RNA.

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ap001

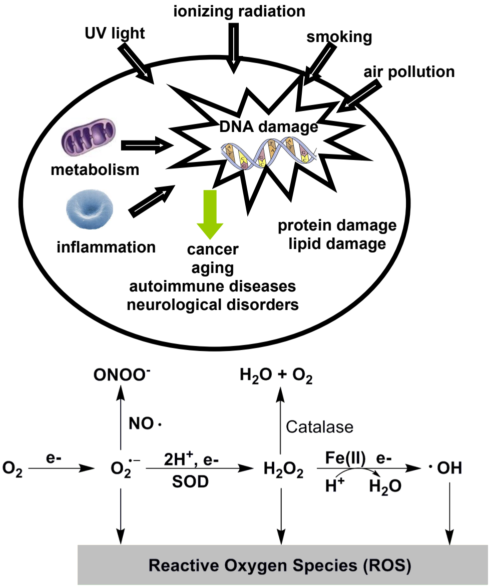

ionizing radiation
UV light
smoking
air pollution
DNA damage
metabolism
inflammation
cancer
aging
autoimmune diseases
neurological disorders
protein damage
lipid damage
ONOO⁻
H₂O + O₂
NO·
Catalase
O₂
e-
O₂·⁻
2H⁺, e-
SOD
H₂O₂
Fe(II)
e-
H⁺
H₂O
·OH
Reactive Oxygen Species (ROS)

Figure 8.1. Reactive oxygen species, endogenously or exogenously produced, are major contributors to DNA damage leading to disease.

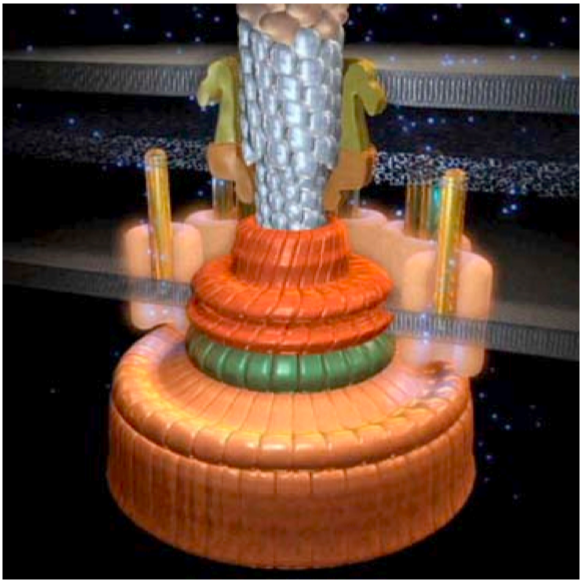


Plate 12.1. Representation of the bacterial flagellar motor. (Reproduced with permission from reference 12.)

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ap001

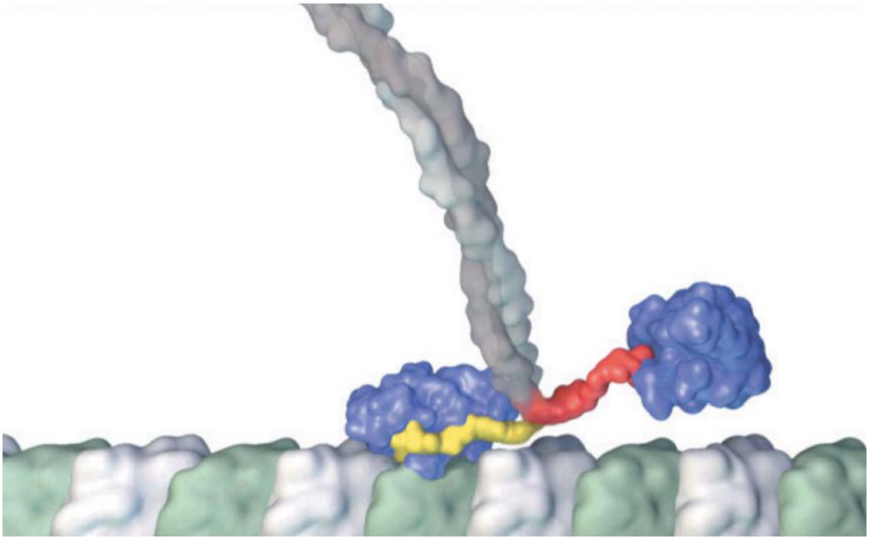


Plate 12.2. Representation of kinesin, a two-headed processive motor, which attaches to and "walks" along a microtubule, shown at bottom. (Partially reproduced from reference 15. Copyright 2000 American Association for the Advancement of Science.)

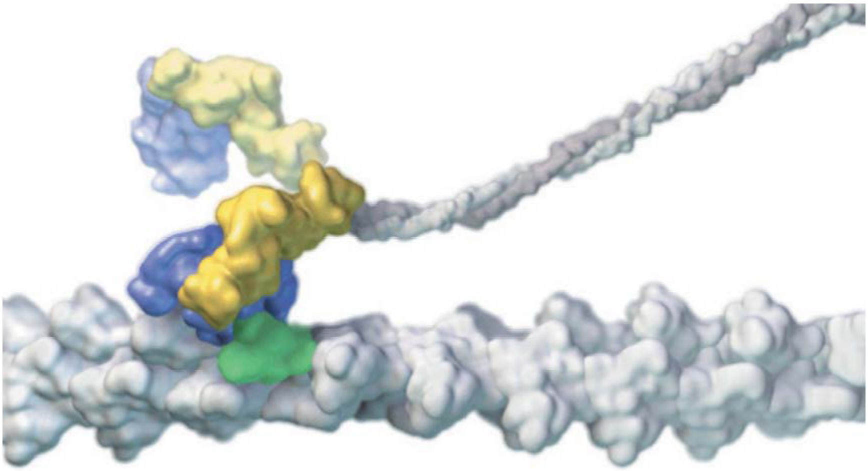


Plate 12.3. Representation of muscle myosin, a two-headed nonprocessive motor, which attaches to actin and delivers one stroke before detaching. (Partially reproduced from reference 15. Copyright 2000 American Association for the Advancement of Science.)

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ap001

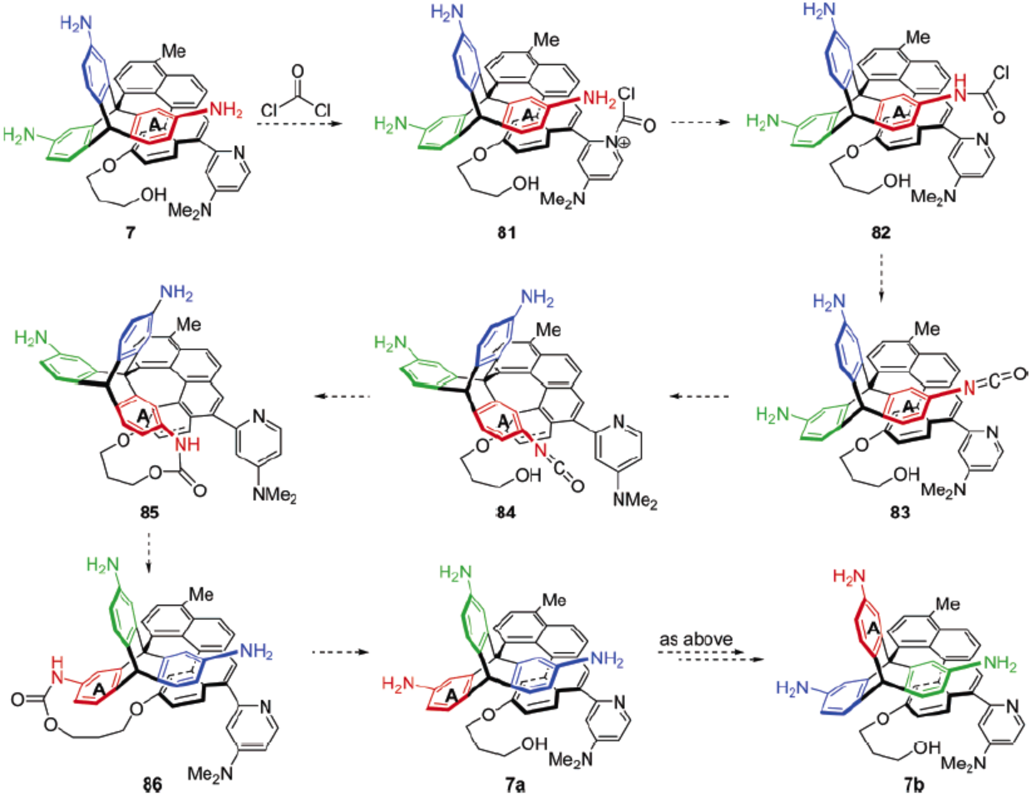

7
81
82
83
84
85
86
7a
7b
as above
A
Me
OH
H_2N
NH_2
Me_2N
NMe_2
Cl
O
N
NH
$N^{\oplus}$

Plate 12.4. Kelly’s chemically-powered ratchet-type rotary motor. (Reproduced from reference 20. Copyright 2007 American Chemical Societv.).

Author Index

Amend, J. P., 63–94

Bean, H. D., 109–132
Bell, T. W., 233–248
Bender, C. J., 313–331
Blackmond, D. G., 133–145
Burrows, C. J., 147–156

Cleaves, H. J., 17–43
Cline, J. I., 233–248

Del Bene, J. E., 95–107

Fujita, E., 283–312

Hall, A. J., 45–62
Hazen, R. M., 3–13
Hemming, N. G., 157–177
Hoye, T. R., 181–203
Hud, N. V., 109–132

Klussman, M., 133–145

Lazcano, A., 17–43
Lynn, D. G., 109–132
McCollom, T. M., 63–94
Martí, A. A., 269–282
Muckerman, J. T., 283–312

Palumbi, S. R., 333–343

Quinn, J. R., 95–107

Rajski, S. R., 205–232
Russell, M. J., 45–62

Seidel, S. R., 249–268
Shavitt, I., 95–107
Shen, B., 205–232
Sizova, E., 181–203

Turro, N. J., 269–282

Zaikowski, L., 249–268
Zimmerman, S. C., 95–107

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix001

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

Subject Index

A

- Acetal-linked nucleic acids (*a*NAs), progenitors of RNA, 126–127
- Acetals, predecessors to phosphodiester bond, 125–127
- Acetate
  - biosynthesis, 51–52
  - carbon oxides and hydrothermal hydrogen synthesizing, 56–57
  - reactions to, 49, 51*f*
- Acetic acid, Kolbe's synthesis, 183
- Acetyl-Co-A pathway
  - biochemical components and energy, 49–50, 51*f*
  - free-energy profiles of acetogenic and methanogenic reduction, 50, 53*f*
- Acidic fluids, radioactive Earth, 46–47, 48*f*
- Acid solutions, metal, phosphate, sulfur, formate, and ammonia sources, 53–55
- Acrolein, precursor of biological amino acids, 25, 26*f*
- Activated acetate, biosynthesis, 51–52
- Adenine
  - anomers of ribosyl nucleosides, 113*f*
  - base pairing, 115
  - base-pairing with thymine, 7
  - small molecules to RNA, 112*f*
  - *See also* Base pairs
- Adenosine triphosphate (ATP)
  - energy, 49, 51*f*
  - pyrophosphate biosynthesis, 50
- Agriculture, use of pesticides, 334–335
- Alanine, Strecker synthesis, 18
- Aldehydes, Miller–Urey experiment, 22, 23*f*
- Aldol reaction, asymmetric, by L-serine, 140–141
- Alizarin, synthetic chemistry, 183*f*
- Alkaline fluids
  - carbon metabolism, 47, 49–50
  - radioactive Earth, 46–47, 48*f*
- Alkaline solutions, metal, phosphate, sulfur, formate, and ammonia sources, 53–55
- Amines
  - calculated Gibbs energies of reaction vs. temperature, 73*f*
  - values of ΔG_i for compounds in energy calculations, 84*t*
  - values of ΔG_r and temperatures for synthesis, 87*t,* 90*t*
- Amino acids
  - acrolein as precursor, 25, 26*f*
  - alkaline conditions, 56–57
  - calculated Gibbs energies of reaction vs. temperature, 73*f*
  - carbon sources, 25, 26*f*
  - catalysis, 140–141
  - "chiral" pairs, 8
  - condensation polymerization, 326–328
  - enantiomeric composition of L-, 139*f*
  - eutectic composition by co-crystallization, 139
  - hydrothermal vents, 35
  - methane as carbon source, 25, 26*f*
  - prebiotic nucleosides and nucleotides, 33–34
  - prebiotic synthesis, 21–25
  - proteinoid material creation, 317, 318*f*
  - stability, 36
  - two-dimensional maps of, melts and proteinoids, 328*f*
  - values of ΔG_i for compounds in energy calculations, 81*t,* 82*t*
  - values of ΔG_r and temperatures for synthesis, 85*t,* 88*t*
  - yields based on initial carbon, 26*f*
- Amino imidazole carbonitrile (AICN), pyrine production, 27–28, 29*f*
- Ammonia
  - acid and alkaline solutions, 53–55
  - Miller–Urey experiment, 22, 23*f*
- Anabolic metabolism, chemical evolution, 315–316
- Ancient ocean
  - applications of boron isotope to, 171–173
  - *See also* Oceans
- Anti-sense region, molecular beacon loop, 271–272
- Apatite, phosphorylating agent, 34
- Aqueous boron species
  - acid-base reaction, 163
  - isotopic composition, 163*f,* 164
  - seawater, 162*f*
  - temperature and salinity, 163*f*
- Arms races, evolutionary, 337–338
- Aromaticity, organic synthesis, 186*f*
- Arsphenamine (Salvarsan), chemical synthesis, 193
- Artificial motors
  - chemically powered, 237–238
  - Kelly's chemically powered ratchet–type rotary motor, 237, 246
- Artificial photosynthesis
  - band-gap-narrowed semiconductors (BGNSCs), 289–292
  - blue dimer catalyst, 293–295
  - catalytic cycle for hydrogen production, 300, 301*f*
  - computational studies for mechanism elucidation, 297–298

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

- directed chemical evolution, 10
- DuBois Ni(II) catalyst, 300, 301*f,* 302
- Fe-only hydrogenase, 300*f*
- H_2 generation and oxidation catalysts by non-noble metal complexes, 299–302
- Ni(diphosphine)$_2$ complexes, 300*f*
- photogeneration of renewable hydride donor, 302–304
- proposed mechanism of water oxidation by Tanaka catalyst, 298, 299*f*
- ruthenium complexes with non-innocent quinone ligands, 295–297
- Ru water oxidation catalysts, 293*f*
- water oxidation catalysis, 293–298

Artificial systems, natural functions, 233–234
Assembly. *See* Coordination-driven self-assembly
Association constants (K_{assoc})
- hydrogen-bonded complexes, 97
- *See also* Base pairs

Asteroids, organic compounds, 37
Asymmetric amplification
- chiral crystals and phase behavior, 137–140
- definition, 135
- enantiomeric composition, 138*f,* 139*f*
- spontaneous synthesis, 136–137

Asymmetric autocatalysis, Soai reaction, 136
Asymmetric synthesis, autocatalysis, 136–137
Atmosphere
- applications of boron isotope to, 171–173
- bacteria, 317
- CO_2 concentrations from past, 158, 159*f*
- Jupiter, 19
- life and, changes affecting climate, 9
- prebiotic synthesis, 21
- primitive terrestrial, 19–21
- primordial, from outgassing, 20

Autocatalysis, asymmetric amplification, 136–137

B

Bacteria
- chemical evolution, 317
- photosynthetic, 64
- thermophilic, in deep-sea vents, 318

Bacterial flagellar motor, evolution product, 235–236, 245
Band-gap-narrowed semiconductors (BGNSCs)
- anion-doped titania, 290
- band gap $E_g(x)$, 291
- capturing photoinduced electrons/holes, 288
- GaN/ZnO solid solutions, 290–292
- in-situ XRD patterns for nitridation of Ga_2O_3/ZnO, 292*f*
- photoelectrolysis, 285, 286*f*
- solar water splitting, 286–287
- variation of band gap vs. ZnO concentration, 292*f*
- visible light absorption and charge separation, 289–292
- *See also* Artificial photosynthesis

Base evolution, proto-RNA polymers, 128, 129
Base excision repair (BER), deoxyribonucleic acid (DNA), 149
Base-pairing
- adenine and thymine, 7
- mispairing, 152, 153*f*
- paradox, 115
- wobble pairs, 152, 153*f*

Base pairs
- association constants (K_{assoc}), 97
- $CH\cdots O$ contacts, 103–104
- classification, 100, 103
- experimental vs. calculated -ΔG for A–T/U and G–C, 102
- hydrogen-bonded structures of A–T, G–C and synthetic, 98*f*
- hydrogen bonding parameters, 97–99
- inosine dimer, 104
- K_{assoc} for complexes of diaminopyridine, 99*f*
- multiple regression analysis, 99–103
- phenomenological analysis, 97
- primary and secondary interactions, 99
- resonance-assisted hydrogen bonds, 100
- single, A–U, 96
- special stability in, 96
- tautomerism, 101
- variables in multivariate regression, 100*t*
- Watson–Crick and Hoogsteen pairing geometries, 103*f*

Base selection, molecular midwives, 117–119, 121–122
Becquerel, Henri, isotope geochemistry, 159
Biaryl lactone rotary motor, chemically powered, 238
Bimolecular complex, templating, 322–323
Binary probes (BPs)
- applications, 275
- before and after target addition, 274*f*
- C-probes, 276*f*
- fluorescence time-decay profiles, 277*f*
- fluorophores, 275
- hybridization probes, 273–274
- oligonucleotide and polynucleotide detection, 276–277
- poly(ethylene glycol)-linked, 276*f*
- probe sequences, 274–275
- pyrene, 276*f*
- ruthenium complex and Cy5 fluorophore, 276*f*
- steady-state luminescence spectra of Ru-Cy5, 278*f*
- target sequence, 275
- three-dye, 276*f*
- time-resolved emission spectra, 278*f*
- two-dye, 276*f*
- *See also* Molecular beacons (MBs)

Bioactive agents, organic synthesis, 193–194

Biochemistry, geochemistry to, 8
Biological homochirality
    definition, 134
    See also Homochirality
Biological systems, artificial systems and natural functions, 233–234
Biological weapons, chemical synthesis, 193
Biology
    ribonucleic acid (RNA), 109–110
    synthesis of oligonucleotides, 122, 123
Biomass
    calculated Gibbs energies of reaction vs. temperature, 73*f*
    composition and synthesis reactions, 77*t*, 78*t*
    energetics of synthesizing monomeric constituents, 74–75
Biomimetic chemistry
    chemical evolution as strategy for synthesis, 315–319
    definition, 314
    natural product synthesis, 195–197
Biomolecules
    acrolein, 25, 26*f*
    ammonia, aldehydes and hydrogen cyanide (HCN) in Miller–Urey experiment, 22, 23*f*
    asteroids, 37
    background research, 18–19
    Bucherer–Bergs synthesis, 23, 24*f*
    Butlerov synthesis, 30–32
    carbohydrates, 30–32
    carbon sources, 25
    comets, 38
    diaminomaleonitrile (DAMN) formation from HCN, 27–28
    energy calculations, 71–72
    extraterrestrial organics, 37–39
    fluxes of HCN to surface of early Earth, 37, 38*f*
    formose reaction, 30, 31*f*
    Haldane, 18
    HCN in prebiotic synthesis, 25
    hydrothermal vents and origins of life, 35–36
    interplanetary dust particles (IDPs), 38–39
    life's chemical origins, 5
    lipids and membrane-forming compounds, 32–33
    methane and amino acid formation, 25, 26*f*
    methane as carbon source, 25, 26*f*
    Miller–Urey experiment, 21, 22*f*
    mirror image "chiral" pairs, 8
    nucleic acid components, 26–32
    Oparin, 18–19
    prebiotic nucleosides and nucleotides, 33–34
    prebiotic synthesis of amino acids, 21–25
    primitive terrestrial atmosphere, 19–21
    purine production from aminoimidazolecarbonitrile (AICN), 29*f*
    purines, 27–28
    pyrimidines, 29–30
    stability at high temperatures, 36
    Strecker synthesis of amino acids, 23, 24*f*
    synthesis of AICN via photoisomerization of DAMN, 28*f*
    Urey, 19
    Wöhler, 18
    See also Energetics of biomolecule synthesis
Biospheres, connection to geospheres, 9
Biosynthesis
    activated acetate, 51–52
    natural products and evolution, 206
    See also Enediyne natural product C-1027; Evolution of organic synthesis
Blue dimer catalyst, water oxidation, 293–295
Bohr, Neil, electrostatic model of atom, 158
Boron isotope pH proxy
    ancient ocean and atmosphere, 171–171
    aqueous species of boron, 163
    aqueous species of boron in seawater, 162*f*
    atmospheric CO_2 concentrations in past, 158, 159*f*
    basic principles, 158
    boron in $CaCO_3$ minerals, 161–162
    boron isotope systematics, 160–164
    Earth history and geological materials, 158
    empirical relationships for seawater/$CaCO_3$ fractionation, 166–168
    ground-truthing studies, 165–168
    ion-exchange apparatus for isolating B from geologic samples, 170*f*
    isotope geochemistry, 159
    isotopic compositions of aqueous boron species, 164
    mass spectrometry analysis, 168
    pH–B isotope relationship, 165–166
    sample preparation, 168–171
Brown hydroboration/oxidation, synthetic methods, 190, 191*f*
Bucherer–Bergs synthesis, amino acids, 23, 24*f*
Building blocks
    calculated Gibbs energies of reaction vs. temperature, 73*f*
    energy calculations, 71–72
    life, 4–6
Bullvalene, degenerate rearrangements, 188, 189*f*
Butlerov synthesis, carbohydrates, 30–32

C

C-1027 natural product. *See* Enediyne natural product C-1027
Calc-alkaline magmatism, emergence of life, 46–47, 48*f*
Calcium carbonate minerals
    boron occurrence, 161–162
    seawater/$CaCO_3$ fractionation, 166–168

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

Calcium phosphate, phosphorylating agent, 34
Camphor, natural product, 194*f*
Carbohydrates
    Butlerov synthesis, 30–32
    formose reaction, 30, 31*f*
Carbon, inventory, 5
Carbon dioxide
    ocean water, 55
    Vostok ice core, record, 172, 173*f*
Carbon metabolism, alkaline solutions toward, 47, 49–50
Carbon sources, amino acids, 25, 26*f*
Catalysts, synthetic methods, 190–192
Catenanes, organic synthesis, 186, 187*f*
Cattle raising, prehistoric, in Europe, 339, 340*f*
Cedrene, natural product, 195*f*
Cell biomass, calculated Gibbs energies of reaction vs. temperature, 73*f*
Cell membranes, synthesis, 32–33
Chemical change. *See* Boron isotope pH proxy
Chemical complexity, surfaces, 7
Chemical environment, Earth's surface, 9
Chemical evolution
    bacteria, 317
    complex evolving systems, 10–11
    critique Harada–Fox polymerization, 321
    description, 314–315
    directed, 10
    interstellar space, 316
    polymerization reactions and Harada–Fox model, 318–319
    primordial Earth conditions, 317
    protein motifs, 325
    spark discharge experiments, 316–317
    templated ligands, 324*f*
    templated reactions and geochemical factor, 320–325
    templating in synthetic chemistry, 322–325
    terrestrial planets, 316
    See also Biomimetic chemistry
Chemical gradients, Earth, 59–60
Chemically powered artificial motors
    chemically powered biaryl lactone rotary motor, 238
    ratchet-type, 237, 246
Chemical origins
    building blocks of life, 4–6
    molecular selection and organization, 6–8
Chemolithoautotrophy, biomass synthesis, 64–65
Chemotrophy, biomass synthesis, 64
Chiral crystals, phase behavior, 137–140
Chirality
    life, 8, 57
    macromolecules, 141–142
    See also Homochirality
"Chiral" pairs, biomolecules, 8
CH···O interactions, inosine dimer pairs, 104
Chromatography, organic synthesis, 184–185
Climate change
    evolutionary marathon, 340–342
    life and atmospheric changes, 9
Cobalt, methyl transfer, 54
Co-evolution, geo- and biospheres, 9
Combinatorial approaches, chemical synthesis, 10
Combinatorial synthesis
    bioactive C-1027 analogs, 226–229
    See also Enediyne natural product C–1027
Comets, organic compounds, 38
Complementary, supramolecular cycles and cages, 250
Complex evolving systems, Earth, 10–11
Condensation polymerization
    modification of Harada–Fox synthesis, 326–327
    products in peptide mixtures, 327–328
Conformational analysis, organic synthesis, 189, 190*f*
Coordination-driven self-assembly
    dodecahedral species, 261–262
    five-by-five methodology, 259–260
    four-by-four methodology, 250–251
    interpenetrated cylinders, 265
    octahedral species, 263
    pentagons and hexagons, 257–258
    pre-designed triangles, 256
    square-triangle equilibria, 253, 254
    supramolecular, star-like pentagon, 258–259
    supramolecular hexagon, 258
    supramolecular hexagons with three-by-three angle-angle methodology, 260–261
    supramolecular pentagon, 257
    supramolecular pentagon by five-by-five methodology, 259*f*
    supramolecular square, 255*f*
    supramolecular squares, 250, 251
    supramolecular triangle, 254*f*
    tetrahedral species, 263
    three-by-three angle-angle methodology, 260–261
    three-dimensional, 261–265
    triangular species with rigid and small subunits, 255–256
    trigonal prism-like species, 264
    two-by-two methodology, 251, 252
    two-by-two "staggered" square, 252–253
    two-dimensional, 250–261
Coordination polymerization, templating, 323
Copper, bioessential element, 9
Corey silyl ether protection, synthetic methods, 190, 191*f*
Cortisone, natural product, 195*f*
Cosmos, origin, 4
Coupling chemistry. *See* Reversible coupling chemistry
C-probes, binary probes, 276*f*
Crown ethers, Williamson synthesis, 322
Crown ligand, encapsulation and specificity, 323, 324*f*

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

Cryptand ligand, encapsulation and specificity, 323, 324*f*
Cyanobacteria, Great Oxygenation Event, 148
Cyanosporasides, biosynthesis, 207, 208*f*
Cyclization
  octatrienes, 189
  template reaction, 323
Cytosine
  anomers of ribosyl nucleosides, 113*f*
  base pairing, 115
  pyrimidine nucleoside formation, 114*f*
  small molecules to RNA, 112*f*

D

Deactivation region, molecular beacon, 272
Degenerate rearrangements, bullvalene, 188, 189*f*
Deoxyribonucleic acid (DNA)
  base excision repair (BER), 149
  biomarker 7,8-dihydro-8-oxo-2'-deoxyguanosine (8-oxoG), 150–151
  damage to human genome, 149–150
  hydantoin products from oxidation of 8-oxoG, 152*f*
  lesions inducible by C-1027, 229, 230*f*
  mispairing, 152, 153*f*
  nucleotide excision repair (NER), 149
  oxidative transformations, 150–151
  pathways to 8-oxoG, 151*f*
  reactive oxygen species, 148*f*
  repair pathways, 149
  wobble pairs, 152, 153*f*
  See also Oxygen atmosphere
Diaminomaleonitrile (DAMN), formation from HCN, 27
Diels–Alder reaction, synthetic methods, 190, 191*f*
DNA. *See* Deoxyribonucleic acid (DNA)
Dodecahedral species, three-dimensional self-assembly, 261–262
Dodecahedrane, natural product, 196*f*
Drosophila melanogaster, genetics, 335, 336f
DuBois Ni(II) catalyst, hydrogen production, 300, 301*f,* 302

E

Earth
  chemical environment of surface, 9
  chemical gradients, 59–60
  complex evolving systems, 10–11
  free oxygen, 20–21
  primordial conditions, 317
Ecstasy, chemical synthesis, 193*f*
Electrically driven, molecular rotor, 240, 241*f*
Electronic theory
  aromaticity, 186*f*
  organic chemistry, 187–189
Elements, inventory for life, 5
Enediyne natural product C-1027
  active site/solvent accessibility of enzyme SgcC4 and tyrosine ammonia lyase (*Rs*TAL), 221–222, 223*f*
  aminomutase SgcC4, 225
  biosynthesis, 207
  biosynthetic gene clusters, 209–212
  biosynthetic pathway β-amino acid component of, 222, 223*f*
  biosynthetic pathway for production of tryptophan, 217*f*
  C-1027, engineered analogs, and biological activities/efficiencies, 229*f*
  combinatorial biosynthesis, 226–229
  cycloaromatization, 207
  discovery of 2-amino-2-deoxyisochorismate (ADIC) synthase activity of SgcD, 219
  deoxyribonucleic acid (DNA) lesions inducible by, 229, 230*f*
  DNA strand cleavage, 209*f*
  domain assignments of enediyne polyketide synthase (PKSEs) from, 214, 215*f*
  enzymes SgcD and SgcG in benzoxazolinate biosynthesis, 217–219
  familial model, 211–212, 213*f,* 214
  formation of 4-methylideneimidazole-5-one (MIO) prosthetic group, 220–224
  gene cluster *sgc*E, 209–210, 212
  isolation from *Streptomyces globisporus*, 207
  mechanistic proposals for MIO-dependent aminomutase chemistry, 225*f*
  metabolic pathways using chorismate as starting material, 217*f*
  minimal PKSE cassettes for C-1027, 212*f*
  neocarzinostatin (NCS) biosynthetic loci, 210–211, 212*f*
  PKSE, 209–216
  production of new bioactive C-1027 analogs, 226–230
  relationship of SgcD, SgcG and SgcCr in, production, 216–225
  role of SgcC4 in β-tyrosine synthesis, 220–225
  Streptomyces genomes, 218
  structure, 208*f*
  views of SgcC4 active site, 226*f*
Enediynes, characterization, 206–207, 208*f*
Energetics of biomolecule synthesis
  alkaline hydrothermal vent fluid mixing with seawater on early Earth, 70*f*
  biomass synthesis reactions, 73, 77*t,* 78*t*
  calculated Gibbs energies of reaction vs. temperature for building blocks of cells, 73*f*
  chemolithoautotrophy, 64–65
  chemotrophy, 64
  composition of late Hadean seawater, 66–67
  composition of late Hadean vent fluid, 68

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

  compositions of end-member fluids in mixing calculations, 67*t*
  energetics (ΔG_r) of synthesis reactions, 72, 79*t*, 80*t*
  energy calculations, 71–72
  mixing calculations, 69–70
  mixing seawater and vent fluid, 65–70
  values of ΔG_i and temperatures for compounds in energy calculations, 81*t*, 82*t*, 83*t*, 84*t*
  values of ΔG_r and temperatures for synthesis reactions, 85*t*, 86*t*, 87*t*, 88*t*, 89*t*, 90*t*
Energy
  acetyl coenzyme-A pathway to acetate, 49, 51*f*
  fossil-based, 284
  water decomposition, 285
Energy calculations, biomolecules, 71–72
Energy consumption, worldwide, 284
Energy sources
  capture, storage and use, 47
  hydrogen, 284–285
  origins-of-life research, 5, 45–46
  requirement for life, 5
Equilenin, natural product, 194*f*
Erythronolide B, natural product, 196*f*
Europe, prehistoric cattle raising, 339, 340*f*
Eutectic point, 138–139
Evolution
  climate change and, 340–342
  complex evolving systems, 10–11
  geo- and biospheres, 9
  natural product biosynthesis, 206
  population, 334
  public health and, 336–337
  RNA world, 128, 129
  secondary metabolism, 206
Evolutionary arms races, 337–338
Evolutionary origins, enediyne polyketide synthase (PKSE), 209–216
Evolution of organic synthesis
  anti-cancer paclitaxel, 193–194
  aromaticity, 186*f*
  bioactive agents, 193–194
  biological weapons, 193*f*
  biomimetic synthesis, 195–197
  bullvalene, 188, 189*f*
  chromatography, 184–185
  conformational analysis, 189, 190*f*
  early history, 182–183
  facile degenerate rearrangements, 188, 189*f*
  ferrocene, 187
  Huang–Minlon modification of Wolff–Kishner reduction, 192*f*
  illicit drugs, 193*f*
  Kolbe's synthesis of acetic acid, 183
  mechanism, 187–189
  natural products, 194–197
  neighboring group participation and proximity acceleration, 188*f*
  Nobel metal-catalyzed synthetic methodologies, 192*f*
  Nobel synthetic methodologies, 191*f*
  octatrienes supporting electrocyclic cyclization, 189*f*
  pharmaceuticals, 193–194
  singlet oxygen ene-reaction, 188*f*
  spectroscopy, 184
  stereochemical course of S_N2 reaction, 188*f*
  strain, 186*f*
  structure, 185–187
  synthetic methodologies, 190–192
  Taxol®, 193–194
  total synthesis publications, 197*t*
  unusual topology, 186, 187*f*
  'what vs. how' of synthesis, 182, 197–198
  See also Natural products
Excimer-monomer switching, molecular beacon, 273*f*
Extraterrestrial organics, flux to Earth, 37–39

F

F_0F_1–ATP synthase, supramolecular motor, 234–235
Fatty acids
  calculated Gibbs energies of reaction vs. temperature, 73*f*
  prebiotic synthesis, 32–33
  values of ΔG_i for compounds in energy calculations, 82*t*, 83*t*
  values of ΔG_r and temperatures for synthesis, 86*t*, 89*t*
Ferrocene, unintended synthesis, 187
Five-by-five methodology, supramolecular pentagon, 259–260
Flavins, redox chemistry, 154
Fluorescent bases, molecular beacon, 273*f*
Fluorescent responsive molecular probes
  oligonucleotides, 10, 270
  See also Binary probes (BPs); Molecular beacons (MBs)
Fluorophores
  binary probes, 275
  molecular beacon reporter, 272
Formate, acid and alkaline solutions, 53–55
Formose reaction
  Butlerov's discovery, 30–31
  sugars on primitive Earth, 31–32
Fossil-based energy, alternatives, 284–285
Fungicides, U.S. agricultural uses, 334, 335*f*

G

Gas chromatography (GC), organic synthesis, 185
Gas–liquid chromatography (GLC), organic synthesis, 185

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

Genetic information. *See* Ribonucleic acid (RNA) world
Genetics, Drosophila melanogaster, 335, 336f
Geochemistry
  isotope, 159
  progression to biochemistry, 8
  proxy, 160
  submarine alkaline vent, 58
  V. M. Goldschmidt, 158
Geology, applications of isotopes in, 160
Geospheres, connection to biospheres, 9
Gibbs energy of reaction (ΔG_r)
  building blocks, 65, 72, 73*f*
  synthesis reactions, 79*t,* 80*t,* 85*t,* 86*t,* 87*t,* 88*t,* 89*t,* 90*t*
Glucose, natural product, 194*f*
Glyoxylate-acetal nucleic acid (*ga*NAs), comparing RNA and, 126–127
Goldschmidt, V. M., father of geochemistry, 158
Great Oxygenation Event
  atmospheric oxygen, 9, 148
  proteins for oxidative stress, 153–154
  See also Oxygen atmosphere
Green rust, stability and Pourbaix diagram, 52*f,* 54, 56*f*
Grignard reagents, synthetic methods, 190, 191*f*
Ground-truthing
  empirical relationships, 166–168
  pH–B isotope relationship, 165–166
  pH control on boron isotopic composition of corals and calcite, 164, 165*f*
  See also Boron isotope pH proxy
Guanine
  anomers of ribosyl nucleosides, 113*f*
  base pairing, 115
  small molecules to RNA, 112*f*

H

Hadean Ocean
  alkaline hydrothermal fluid entering, 50, 52*f*
  dissolved CO_2 and life, 60
  emergence of life, 47, 48*f,* 49*f*
  ferrous ions, 53
Haemin, natural product, 194*f*
Haldane, J. B. S., hypothesis, 18
Harada–Fox model
  critique, 321
  modification, 326–327
  polymerization reactions, 318–319
Herbicides, U.S. agricultural uses, 334, 335*f*
Heterochiral peptides, synthesis, 56–57
Histidine, enantiomeric composition, 139*f*
Homochirality
  amino acid catalysis, 140–141
  asymmetric aldol reaction by L-serine, 141*f*
  asymmetric amplification, 136–140
  biological, 134
  chiral crystals and phase behavior, 137–140
  chirality in macromolecules, 141–142
  chirality transfer and preservation, 140–142
  chiroselective self-replication of homochiral peptide LL, 142*f*
  crystal forms of chiral substances, 138*f*
  definition, 134
  enantiomeric composition of L-amino acids, 139*f*
  enhancing eutectic composition of valine, 139*f*
  importance, 134–135
  mirror symmetry breaking, 135–136
  solid-solution equilibrium of enantiomerically enriched substance, 138*f*
  spontaneous asymmetric synthesis, 136–137
Homosecodaphniphyllate, biomimetic synthesis, 195, 197*f*
Hoogsteen pairing, geometry for A–T, 101, 103*f*
Huang–Minlon modification, Wolff–Kishner reduction, 192
Human genome
  daily damage, 149–150
  oxidative stress, 148
Hydride donor, photogeneration of renewable, 302–304
Hydrogen
  catalytic production using DuBois Ni(II) catalyst, 300, 301*f,* 302
  energy source, 284–285
  generation by non-noble metal complexes, 299–302
  inventory, 5
  photoelectrolysis, 285–286
  See also Artificial photosynthesis
Hydrogen, dissolved, reducing CO_2, 5
Hydrogen bonding
  diaminopyridine and thymine, 97, 99*f*
  multiple regression analysis, 99–103
  parameters, 97–99
  primary and secondary interactions, 99
  proton acidity, 97
  resonance-assisted, 100
  structures of base pairs, 98*f*
Hydrogen cyanide (HCN)
  flux to Earth, 37, 38*f*
  intermediate in prebiotic synthesis, 25
  Miller–Urey experiment, 22, 23*f*
Hydrothermal vents
  mixing seawater and vent fluid, 65–70
  origin of life, 35–36
  submarine, 36

I

Illicit drugs, chemical synthesis, 193
Imine/amine chemistry, replication, 124–125
Indigo, synthetic chemistry, 183*f*

Inorganic compounds, values of ΔG_i for compounds in energy calculations, 81*t*
Inosine dimmer, complex, 104
Insecticides, U.S. agricultural uses, 334, 335*f*
Intercalators
  acting as midwives, 121*f*
  template directed ligation reaction, 121–122
Interpenetrated cylinders, three-dimensional self-assembly, 265
Interplanetary dust particles (IDPs), organic compounds, 38–39
Interstellar molecules, organic matter, 37
Interweaving, template reaction, 323
Ion-exchange apparatus, boron from geologic samples, 170–171
Iron
  bioessential element, 9
  ocean, 54
Iron-nickel sulfide minerals, biomolecule catalysts, 5
Isoleucine, enantiomeric composition, 139*f*
Isotopes. *See* Boron isotope pH proxy
Isovaline, catalysis, 140–141

J

Jupiter, atmosphere, 19

K

Keeling, D., atmospheric CO_2 concentrations, 158
Kinesins, transport and muscle proteins, 236, 245
Kinetic barrier, methane synthesis, 49
Knowles asymmetric reduction, metal-catalyzed synthetic methodologies, 191, 192*f*
Kolbe's synthesis, acetic acid, 183

L

Lactase, dairy diet, 339
Late Hadean era, life emergence on Earth, 64
Late Hadean seawater, composition, 66–67
Late Hadean vent fluid, composition, 68
Leucine, enantiomeric composition, 139*f*
Life
  atmospheric changes and, affecting climate, 9
  building blocks of, 4–6
  chirality of, 8, 57
  complex evolving systems, 10–11
  emergence in late Hadean era, 64
  energy capture, storage and use, 47
  evolution story, 45–46
  hydrothermal vents and origin of, 35–36
  initial conditions for emergence, 46, 48*f*
  time window for origin of, 19–20
Light-powered molecular motors
  artificial, 238–239
  chiral imine, 239, 240*f*
Linear, template reaction, 323
Linkage evolution, proto-RNA polymers, 128, 129
Lipids, prebiotic synthesis, 32–33
Liquid chromatography–mass spectrometry (GC–MS), organic synthesis, 185
Loop, molecular beacon, 271–272
Lost City, submarine alkaline vent, 58

M

Mackinawite, stability and Pourbaix diagram, 54, 56*f*
Macromolecules
  chirality, 141–142
  self-assembly, 7
Magnesium-iron silicates, oceanic crust, 46–47, 48*f*
Mannich reactions, autocatalysis, 137
Marine hydrothermal systems, origin of life, 65
Mauna Loa, atmospheric CO_2 concentrations, 158, 159*f*
Mauveines, synthetic chemistry, 183*f*
Mechanisms, organic synthesis, 187–189
Membrane-forming compounds, prebiotic synthesis, 32–33
Merrifield solid phase synthesis, synthetic methods, 190, 191*f*
Metabolism, chemical gradients and Earth, 59–60
Metal ions, condensation polymerization, 326–328
Meteorites, discovery of organic compounds, 5
Methamphetamine, chemical synthesis, 193*f*
Methane, source of amino acids, 25, 26*f*
Methicillin resistant *Staphylococcus aureus* (MRSA) infection, 337–338
Methionine, enantiomeric composition, 139*f*
Microbiology, modern submarine alkaline vent, 58
Miller–Urey experiment
  ammonia, aldehydes and hydrogen cyanide (HCN), 22, 23*f*
  organic compounds, 21, 22*f*
Mirror image "chiral" pairs, biomolecules, 8
Mirror symmetry, universe, 135–136
Mixing calculations, seawater and hydrothermal fluids, 67*t*, 69–70
Möbius topology, organic synthesis, 186, 187*f*
Molecular beacons (MBs)
  anti-sense region, 271–272
  "closed" conformation, 270, 271*f*
  complementary or anti-sense sequence, 272
  "deactivation region," 272
  excimer-monomer switching, 273*f*
  MB reporter, 272

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

  - MBs with fluorescent bases, 273*f*
  - multiple-pyrene labeled, 273*f*
  - "open" conformation, 270, 271*f*
  - quadruplex-based, 273*f*
  - self-complementary region (stem), 271
  - steady-state fluorescence spectra, 271*f*
  - three-dye, 273*f*
  - triplex, 273*f*
  - two-dye, 273*f*
  - *See also* Binary probes (BPs)
- Molecular biology, ribonucleic acid (RNA), 109–110
- Molecular imprinting
  - chemical evolution of protein motifs, 325
  - templating, 323–324
- Molecular midwives
  - base pairing, 118
  - base selection and preorganization, 117–119, 121–122
  - hypothesis, 119*f*
  - intercalators acting as, 121*f*
  - nucleoside base assemblies and, 120*f*
- Molecular motors
  - bacterial flagellar motor, 235–236, 245
  - Branchaud's chemically powered biaryl lactone rotary, 238*f*
  - chemically powered artificial, 237–238
  - designed, 236–244
  - electrically driven molecular rotor, 240, 241*f*
  - F_0F_1–ATP synthase, 234–235
  - Feringa's second-generation light-powered, 239*f*
  - Kelly's chemically powered ratchet-type rotary motor, 246
  - Lehn's putative light-powered chiral imine, 240*f*
  - light-powered, 238–239
  - natural, 234–236
  - photoisomerization quantum efficiency of 2-substituted 9-(2,2,2-triphenylethylidene)fluorenes, 244*t*
  - photon-driven ratchet design, 241–244
  - Rapenne's electrically driven molecular rotor, 240, 241*f*
  - retrosynthesis of target molecular motor, 242*f*
  - synthesis of (*E*)- and (*Z*)-2-*t*-butyl-9-(2,2,2-triphenylethylidene)fluorenes, 243*f*
  - synthesis of model rotors with nitrogen substituents, 243*f*
  - transport and muscle proteins, 236
  - two-headed nonprocessive motor, 245
  - two-headed processive motor, 245
  - work, 234
- Molecular probes. *See* Binary probes (BPs); Molecular beacons (MBs)
- Molecular rotor, electrically driven, 240, 241*f*
- Molybdenum
  - alkaline fluids, 54
  - bioessential element, 9
- Morphine, natural product, 195*f*
- Multiple-pyrene labeled, molecular beacon, 273*f*
- Multiple regression analysis, hydrogen bonding, 99–103
- Muscle proteins, transport and, 236
- Mustard gas, chemical synthesis, 193*f*
- Mutagenicity, deoxyribonucleic acid (DNA), 153
- Myosin, transport and muscle proteins, 236, 245

N

- Natural products
  - artificial systems and natural functions, 233–234
  - biomimetic synthesis, 195–197
  - chemical synthesis, 194–197
  - combinatorial biosynthesis, 226–229
  - dodecahedrane, 196*f*
  - erythronolide B, 196*f*
  - evolution, 206
  - homosecodaphniphyllate, 197*f*
  - palytoxin, 196*f*
  - progesterone, 195, 197*f*
  - targets from 1950s, 195*f*
  - "total synthesis" publications, 197*t*
  - vitamin B_{12}, 196*f*
  - *See also* Enediyne natural product C-1027; Evolution of organic synthesis
- Natural selection
  - evolution writ large, 339
  - population evolution, 334
- Neighboring group participation, studies, 188
- Neocarzinostatin, enediyne cyclization, 207, 208*f*
- Neoprene, discovery, 183
- NH···N hydrogen bond, inosine dimer pairs, 104
- NH···O hydrogen bonds, inosine dimer pairs, 104
- Nickel, catalyst, 54
- Nitrogen
  - inventory, 5
  - ocean water, 55
  - Pourbaix diagram and stability, 58*f*
- Nobel
  - metal-catalyzed synthetic methodologies, 192*f*
  - synthetic methodologies, 190, 191*f*
- Non-enzymatic template directed replication, nucleic acids, 116–117
- Nonprocessive motor, two-headed, 236, 245
- Noyori asymmetric reduction, metal-catalyzed synthetic methodologies, 191, 192*f*
- Nucleic acid bases, base pairing, 115
- Nucleic acid components
  - carbohydrates, 30–32
  - purines, 27–28, 29*f*
  - pyrimidines, 29–30

Nucleic acids, non-enzymatic template directed replication, 116–117
Nucleosides
  coupling bases with ribose, 113–114
  formation, 111, 113–114
  prebiotic, 33–34
Nucleotide excision repair (NER), deoxyribonucleic acid (DNA), 149
Nucleotides
  calculated Gibbs energies of reaction vs. temperature, 73*f*
  prebiotic, 33–34
  synthesis, 115–117
  values of ΔG_i for compounds in energy calculations, 82*t*
  values of ΔG_r and temperatures for synthesis, 85*t*, 86*t*, 89*t*
Nylon, "cold drawing," 183

O

Obduction zones, emergence of life, 48*f*
Oceans
  applications of boron isotope to ancient, 171–173
  bacteria, 317
  prebiotic organic synthesis, 6
  sediment core, 173*f*
  surface, 7
Octahedral species, three-dimensional self-assembly, 262, 263
Octatrienes, cyclization, 189
Oligonucleotides
  binary probe case study, 276–277
  binary probes, 273–277
  biological synthesis, 122, 123
  fluorescent responsive molecular probes, 10, 270
  molecular beacons, 270–272, 273*f*
  synthesis, 115–117
  See also Binary probes (BPs); Molecular beacons (MBs)
Oparin, A. I., origin of life, 18–19
Organic chemistry, mechanism, 187–189
Organic compounds, extraterrestrial contribution, 37–39
Organic molecules
  discovery in carbonaceous meteorites, 5
  selection and organization, 6–8
  See also Evolution of organic synthesis
Organization, organic molecules, 6–8
Origin of life
  autogenic theories, 59
  chirality, 8, 57
  energy, 5, 45–46
  hydrothermal vents and, 35–36
  marine hydrothermal systems, 65
  primitive terrestrial atmosphere, 19–21
  See also Ribonucleic acid (RNA) world
Outgassing
  primordial atmosphere, 20
  Rubey model, 19
Oxidation catalysis, hydrogen generation and, by non-noble metal complexes, 299–302
Oxidative stress, cell's genome, 148
Oxygen
  free, in early Earth, 20–21
  increments in atmospheric, 9
  inventory, 5
Oxygen atmosphere
  analogy between redox chemistry of 7,8-dihydro-8-oxo-2'-deoxyguanosine (8-oxoG) and pterins and flavins, 154
  base excision repair (BER) pathway, 149
  biomarker 8-oxoG for oxidative stress, 150–151
  damage to human genome, 149–150
  DNA damage and repair, 155
  DNA lesions and frequency of occurrence, 150*f*
  DNA repair pathways, 149*f*
  formation of hydantoin products from oxidation of 8-oxoG, 151, 152*f*
  G-to-C mutations, 152, 153*f*
  mechanism of G oxidation to form hydantoin lesions, 151
  mispair of G-to-T transversion mutations via 8-oxoG, 152, 153*f*
  mutagenicity, 153
  mutation and lifespan, 153–154
  nucleotide excision repair (NER), 149*f*
  8-oxoG pathways, 150–151
  wobble pairs, 152, 153*f*

P

Paclitaxel, chemical synthesis, 193–194
Palytoxin, natural product, 196*f*
Penicillin V, natural product, 195*f*
Peptide nucleic acid (PNA), prebiotic synthesis, 34
Peptides
  heterochiral, 56–57
  pH and formation of phenylalanine dipeptide, 57, 59*f*
  self-replication of homochiral, 142
Pesticides, agricultural uses, 334–335
Pharmaceuticals, organic synthesis, 193–194
Phase behavior, chiral crystals, 137–140
Phosphate, acid and alkaline solutions, 53–55
Phosphodiester formation
  acetals as predecessors to, 125–127
  nucleotide and oligonucleotide, 115–116
Phospholipid molecules, self-organization, 7
Phosphorus, inventory, 5
Phosphorylation, nucleosides, 33–34
Photoelectrolysis
  band-gap-narrowed semiconductors (BGNSCs), 286–288

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

  discovery using *n*-type TiO_2, 286–287
  See also Artificial photosynthesis
Photogeneration, renewable hydride donor, 302–304
Photon-driven ratchet, motor design, 241–244
Photosynthesis
  artificial, 10
  diagram of natural, 288*f*
  oxygenic, 64
  See also Artificial photosynthesis
pH proxy. *See* Boron isotope pH proxy
Planet accretion, temperature, 20
Podand ligand, encapsulation and specificity, 323, 324*f*
Polyether ring, templating, 323
Poly(ethylene glycol), binary probes, 276*f*
Polyketide synthases enediyne (PKSE), evolutionary origins, 209–216
Polymerization reactions, Harada–Fox model, 318–319
Polynucleotide detection, binary probes, 276–277
Populations, evolution, 334
Post-glacial Neoproterozoic events, atmospheric oxygen, 9
Pourbaix diagram
  nitrogen species, 58*f*
  pH boundary of
    monophosphate/polyphosphate, 50, 52*f*
  siderite, mackinawite, and green rust, 54, 56*f*
Prebiotic organic synthesis
  amino acids, 21–25
  oceans, 6
  pathways to sugar, 6–7
  reducing atmosphere, 21
  ribonucleic acid (RNA), 110
Pre-designed triangle, two-dimensional assembly, 256
Primitive terrestrial atmosphere, origin of life, 19–21
Primordial atmosphere, outgassing, 20
Probe sequences, binary probes, 274–275
Processive motor, two-headed, 236, 245
Progesterone, biomimetic synthesis, 195, 197*f*
Prontosil, chemical synthesis, 193
Proteinoid material, amino acids in hot crucible, 317, 318*f*
Proteins
  chemical evolution, 315
  resolution and discrimination of products in
    peptide mixtures, 327–328
  templating and chemical evolution, 325
  See also Biomimetic chemistry
Proximity acceleration, studies, 188
Proxy
  geochemistry, 160
  See also Boron isotope pH proxy
Pterins, redox chemistry, 154
Publications, "total synthesis," 197*t*
Public health, and evolution, 336–337
Purines
  hydrothermal vents, 35
  prebiotic nucleosides and nucleotides, 33–34
  prebiotic synthesis, 27–28
  production from amino imidazole
    carbonitrile (AICN), 29*f*
  stability, 36
Pyrene, binary probes, 276*f*
Pyrimidines
  hydrothermal vents, 35
  possible mechanisms for prebiotic synthesis, 30*f*
  prebiotic nucleosides and nucleotides, 33
  prebiotic synthesis, 29–30
  stability, 36
2-Pyrimidinone, pyrimidine nucleoside formation, 114*f*
Pyrophosphate, biosynthesis, 50, 52*f*

Q

Quadruplex, molecular beacon, 273*f*
Quencher, molecular beacon, 272
Quinine, natural product, 194*f*

R

Racemic compounds, 138
Radioactive Earth, acidic and alkaline fluids, 46–47
Ratchet motor design, photon-driven, 241–244
Rayon, discovery, 183
Reactive oxygen species (ROS), deoxyribonucleic acid (DNA), 148
Reagents, synthetic methods, 190–192
Redox chemistry, pterins and flavins, 154
Redox reactions, early life, 5–6
Reducing atmosphere, prebiotic synthesis, 21
Reduction, Huang–Minlon modification of Wolff–Kishner, 192
Reductive amination, reversible backbone linkages, 123–125
Renewable hydride donor, photogeneration, 302–304
Reporter, molecular beacon, 272
Reserpine, natural product, 195*f*
Resonance-assisted hydrogen bonding, theory, 100
Reversible coupling chemistry
  acetal-linked nucleic acids (*a*NAs), 126–127
  acetals as predecessors to phosphodiester
    bond, 125–127
  bifunctional amine-nucleoside monomers,
    124, 125*f*
  biological synthesis of oligonucleotides,
    122, 123
  catalytic activity of DNA template, 123,
    124*f*

  comparing RNA and glyoxylate-acetal nucleic acid, 126–127
  imine/amine system, 124–125
  product distribution for reductive polymerization, 126*f*
  reductive amination, 123–125
  thermodynamically controlled polymerization, 122–127
Ribonucleic acid (RNA) world
  acetals as predecessors to phosphodiester bond, 125–127
  anomers of ribosyl nucleosides, 113*f*
  AU base assemblies, 120*f*
  base selection and preorganization, 117–119, 121–122
  bifunctional amine-nucleoside monomers in replication of DNA templates, 125*f*
  biological synthesis of oligonucleotides, 123*f*
  bridge to origin of life, 109–110
  catalytic activity of simple DNA template, 124*f*
  coupling and nucleoside problem, 113–114
  evolution and emergence of, 128
  gap in prebiotic timeline, 111–117
  GC base assemblies, 120*f*
  hypothesis, 154
  intercalators as midwives in template directed ligation reaction, 121*f*
  linkage comparison of RNA and glyoxylate-acetal nucleic acid, 127*f*
  missing link, 111–117
  'molecular midwife' hypothesis, 119*f*
  molecular midwives, 117–119, 121–122
  nucleoside base assemblies and molecular midwives, 120*f*
  nucleoside formation, 111, 113–114
  nucleotide and oligonucleotide synthesis, 115–117
  paradox of base pairing, 115
  product distribution upon reductive polymerization of monomer T_1, 126*f*
  proposed mechanism of pyrimidine nucleoside formation, 114*f*
  proto-RNA polymers, 128, 129
  reductive amination for reversible backbone linkages, 123–125
  small molecule world to, 112*f*
Ribose
  anomers of ribosyl nucleosides, 113*f*
  small molecules to RNA, 112*f*
Ribose synthesis, prebiotic, 34
RNA. *See* Ribonucleic acid (RNA) world
Robinson annulation, synthetic methods, 190, 191*f*
Rocks, surfaces, 7
Rotary motor
  chemically powered ratchet–type, 237, 246
  See also Molecular motors
Rotaxane, organic synthesis, 186, 187*f*
Rotors
  electrically driven molecular rotor, 240, 241*f*
  synthesis of model, with nitrogen substituents, 243*f*, 244
  See also Molecular motors
Rubey, outgassing model, 19
Ruthenium catalyst
  renewable hydride donor, 302–304
  water oxidation, 295–297
Ruthenium complex with fluorophore
  binary probes, 276*f*
  fluorescence time-decay profile, 277*f*
  steady-state luminescence, 278*f*
Rutherford, "planetary" model, 158

S

Sabatier process, metal-catalyzed synthetic methodologies, 191, 192*f*
Saccharides
  calculated Gibbs energies of reaction vs. temperature, 73*f*
  values of ΔG_i for compounds in energy calculations, 83*t*
  values of ΔG_r and temperatures for synthesis, 87*t*, 90*t*
Salvarsan, chemical synthesis, 193
Sarin, chemical synthesis, 193*f*
Seawater
  aqueous species of boron, 162*f*
  composition of fluid during mixing with vent fluid, 70*f*
  composition of late Hadean, 66–67
  continental ice volume signal from $\delta^{18}O$, 172, 173*f*
  mixing, and vent fluid, 65–70
  mixing calculations, 69–70
  polymerization reactions, 318–319
Secondary metabolism, evolution, 206
Selection, organic molecules, 6–8
Self-assembly
  life on Earth, 249
  RNA-like polymers, 7, 109–110
  See also Coordination-driven self-assembly; Ribonucleic acid (RNA) world
Self-complementary region, molecular beacon stem, 271
Self-correction, phenomenon, 249–250
Self-organization, phospholipids, 7
Self-replication, homochiral peptides, 142
Serine
  asymmetric aldol reaction by L-, 140–141
  enantiomeric composition, 139*f*
Serpentinization, oceanic crust, 46–47
Sharpless asymmetric dihydroxylation, metal-catalyzed synthetic methodologies, 191, 192*f*

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002

Sharpless asymmetric epoxidation, metal-catalyzed synthetic methodologies, 191, 192*f*
Siderite, stability and Pourbaix diagram, 52*f*, 53, 56*f*
Singlet oxygen ene-reaction, mechanism, 188
"Snowball Earth" period, theory, 148
Soai reaction, asymmetric autocatalysis, 136–137
Solar energy, 284, 302–304
Spark discharge, chemical evolution, 316–317
Spectroscopy, organic synthesis, 184
Spherand ligand, encapsulation and specificity, 323, 324*f*
Spontaneous asymmetric synthesis, autocatalysis, 136–137
Sporolides, biosynthesis, 207, 208*f*
Stability
  base pairs, 96
  biomolecules at high temperature, 36
Stacking, base pairing, 115
Staphylococcus aureus, history of antibiotic treatment, 337, 338*t*
Stem, molecular beacon, 271
Stereochemical course, S_N2 reaction, 188
Stoichiometric reagents, synthetic methods, 190, 191*f*
Strain, organic synthesis, 186*f*
Strecker, A., alanine synthesis, 18
Strecker synthesis, amino acids, 23, 24*f*
Streptomyces globisporus, isolation of enediyne C-1027, 207
Structure, organic synthesis, 185–187
Subduction zones, emergence of life, 48*f*
Submarine vents
  microbiology of modern, 58
  organic compound destruction, 36
  See also Vents
Sugar evolution, proto-RNA polymers, 128, 129
Sugar molecules
  "chiral" pairs, 8
  prebiotic synthesis pathways, 6–7
  stability, 36
Sugar phosphate polymers, prebiotic synthesis, 34
Sulfur
  acid and alkaline solutions, 53–55
  inventory, 5
Sun
  energy source for life, 5
  solar energy, 284, 302–304
Supramolecular hexagons
  three-by-three angle-angle methodology, 260–261
  two-dimensional assembly, 257–258
Supramolecular motor
  F_0F_1–ATP synthase, 234–235
  See also Molecular motors
Supramolecular pentagons
  five-by-five methodology, 259
  star-like, 258–259
  two-dimensional assembly, 257
Supramolecular squares
  equilibrium between square and triangle, 253, 254
  four-by-four methodology, 250–251
  two-by-two methodology, 251, 252
  two-by-two "staggered" square, 252–253
  two-dimensional assembly, 255
Supramolecular triangle
  equilibrium between square and triangle, 253, 254
  two-dimensional assembly, 254
Surfaces, chemical complexity, 7
Symmetry breaking, definition, 135
Synthesis
  definition, 182
  publications with "total synthesis," 197*t*
  'what vs. how,' 182, 197–198
  See also Evolution of organic synthesis
Synthetic chemistry
  templating, 322–325
  Williamson synthesis, 322
Synthetic methodologies, reagents and catalysts, 190–192

T

Tanaka catalyst
  computational studies for elucidation of mechanism, 297–298
  proposed mechanism of water oxidation, 298, 299*f*
  ruthenium complexes, 295–297
  structure, 296*f*
Target
  binary probes before and after, 274*f*
  molecular beacon, 272
  sequence for binary probes, 275
Tautomerism, base pairs, 101
Taxol®, chemical synthesis, 193–194
Teflon®, polymerization, 183
Temperature
  planet accretion, 20
  stability of biomolecules at high, 36
Templating
  chemical evolution of protein motifs, 325
  ligands increasing level of encapsulation and specificity, 324*f*
  molecular imprinting, 323–324
α-Terpineol, natural product, 194*f*
Terrestrial atmosphere, primitive, 19–21
Terrestrial planets, chemical evolution, 316
Tetrahedral species, three-dimensional self-assembly, 263
Tetrahydrogestrinone (THG), chemical synthesis, 193*f*
Thin-layer chromatography (TLC), organic synthesis, 185
Tholins, organic polymers, 39

Three-by-three angle-angle methodology, supramolecular hexagons, 260–261
Three-dye
  binary probes, 276*f*
  molecular beacon, 273*f*
Thymine
  base-pairing with adenine, 7
  See also Base pairs
"Topicity," supramolecular cycles and cages, 250
Topology, organic synthesis, 186, 187*f*
Transport, muscle proteins, 236
Triangles
  predesigned, by two-dimensional assembly, 256
  rigid and small subunits, 255–256
  supramolecular, 254, 255
Trigonal prism-like species, three-dimensional self-assembly, 264
Triplex, molecular beacon, 273*f*
Tropinone, natural product, 194*f*
Tungsten, alkaline fluids, 54
Two-by-two methodology, supramolecular squares, 251, 252
Two-by-two "staggered" square, two-dimensional assembly, 252–253
Two-dye
  binary probes, 276*f*
  molecular beacon, 273*f*
Two-headed nonprocessive motor, 236, 245
Two-headed processive motor, 236, 245

U

Uracil
  anomers of ribosyl nucleosides, 113*f*
  base pairing, 115
  pyrimidine nucleoside formation, 114*f*
  single base pair A–U, 96
  small molecules to RNA, 112*f*
  See also Base pairs
Urea
  synthesis by Wöhler, 18
  synthetic chemistry, 182, 183*f*

V

Valine, enhancing eutectic composition, 139*f*
Vents
  composition of, fluid mixing with seawater, 70*f*
  composition of late Hadean fluid, 68
  hydrothermal, and origin of life, 35–36
  microbiology of modern submarine alkaline, 58
  seawater at deep-sea thermal, 318–319
  thermophilic bacteria, 318
Vitamin B_{12}, natural product, 196*f*
Volcanism, emergence of life, 46–47, 48*f*
Vostok ice core, CO_2 record, 173*f*
Vulcanization of rubber, accidental, 183

W

Water, building block of life, 4–5
Water decomposition, hydrogen generation, 284–285
Water oxidation
  base titration of Ru complexes, 296–297
  blue dimer catalyst, 293–295
  catalysts, 293
  computational studies of mechanism, 297–298
  mechanisms for oxygen evolution from blue dimer, 294*f*
  proposed mechanism by Tanaka catalyst, 298, 299*f*
  ruthenium complexes with non-innocent quinone ligands, 295–297
  Tanaka catalyst and acid-base reaction, 296*f*
  See also Artificial photosynthesis
Watson–Crick pairing
  base pairing, 115
  geometry for A–T, 101, 103*f*
Wavelength shifting, molecular beacon, 273*f*
Williamson synthesis, crown ethers, 322
Wittig olefination, synthetic methods, 190, 191*f*
Wöhler, F., urea synthesis, 18
Wolff–Kishner reduction, Huang–Minlon modification, 192
Work
  definition, 234
  See also Molecular motors

Downloaded by 89.163.35.42 on June 27, 2012 | http://pubs.acs.org
Publication Date (Web): December 20, 2009 | doi: 10.1021/bk-2009-1025.ix002